A DICTIONARY

OF

APPLIED PHYSICS

MACMILLAN AND CO., Limited
LONDON • BOMBAY • CALCUTTA • MADRAS
MELBOURNE

THE MACMILLAN COMPANY
NEW YORK • BOSTON • CHICAGO
DALLAS • SAN FRANCISCO

THE MACMILLAN CO. OF CANADA, Ltd.
TORONTO

A DICTIONARY

OF

APPLIED PHYSICS

EDITED BY

SIR RICHARD GLAZEBROOK

K.C.B., D.Sc., F.R.S.

IN FIVE VOLUMES

VOL. IV

LIGHT—SOUND—RADIOLOGY

MACMILLAN AND CO., LIMITED
ST. MARTIN'S STREET, LONDON
1923

PRINTED IN GREAT BRITAIN

LIST OF CONTRIBUTORS

LIGHT—SOUND—RADIOLOGY

ALDIS, A. CYRIL W., M.A., M.B.E. . . . . . A. C. W. A.
Projection Apparatus.

ANDERSON, JOHN A., Ph.D., Mt. Wilson Observatory, U.S.A. . . . J. A. A.
Diffraction Gratings, Manufacture and Testing.

ANDERSON, JOHN S., M.A., D.Sc., Ph.D. . . . . . J. S. A.
Camera Lenses, Testing of.
Immersion Refractometry.
Light, Interference of.
Objectives, Testing of Compound.
Shutters, Testing of Photographic.
Telescopes of German Design.
Ultramicroscope.

BAKER, THOS. Y., Instructor-Commander, R.N., B.A., F.Inst.P. . . T. Y. B.
Navigation and Navigational Instruments.

BARNARD, JOHN EDWIN, F.Inst.P. . . . . . . J. E. B.
Microscopy with Ultra-violet Light.

BARTON, E. H., D.Sc., F.R.S., F.Inst.P. . . . . . E. H. B.
Sound.

BRAGG, WILLIAM LAWRENCE, M.A., O.B.E., M.C. . . . . W. L. B.
Sound Ranging.

CHESHIRE, FREDERIC J., C.B.E., F.Inst.P. . . . . . F. J. C.
Range-finders, Short-base.

CLAY, REGINALD S., B.A., D.Sc., F.Inst.P. . . . . . R. S. C.
Phonograph and Gramophone.
Pianoforte.
Piano-player.

COAD-PRYOR, E. A., B.A. . . . . . . E. A. C.-P.
Glass.

COBLENTZ, W. W., Ph.D., Bureau of Standards, U.S.A. . . . . W. W. C.
Infra-red Transmission and Refraction Data.
Radiation, Determination of Constants of.

CONRADY, A. E., Professor of Optical Design, Imperial College . . . A. E. C.
Microscope, Optics of the.

DALE, Miss A. BARBARA, M.A. . . . . . . A. B. D.
Lenses, Testing of Simple.
Polarised Light.

DARWIN, C. G., M.A., F.R.S. . . . . C. G. D.
Quantum Theory.
Radiation Theory.

FRENCH, JAMES WEIR, D.Sc., F.Inst.P. . . . . . J. W. F.
Kinematograph, The.
Optical Parts, The Working of.

GAMBLE, CHAS. W., M.Sc.Tech., O.B.E. (Military) . . . . . C. W. G.
Photographic Apparatus.

GIBSON, K. S., Ph.D., Bureau of Standards, U.S.A. . . . . K. S. G.
Spectrophotometry.

GUILD, J., A.R.C.Sc., D.I.C., F.Inst.P. . . . . . . J. G.
Diffraction Gratings Theory.
Eye, The.
Goniometry.
Interferometers : Technical Applications.
Spectroscopes and Refractometers.
Spherometry.

HENRICI, E. O., Major R.E. (ret.), F.R.A.S., F.Inst.P., Associate Inst.C.E. . E. O. H.
Divided Circles (with G. W. Watts).
Spirit-levels.
Surveying and Surveying Instruments.
Theodolite, The.

HILL, A. V., Sc.D., F.R.S. . . . . A. V. H.
Position-finder, The Mirror.

KAYE, G. W. C., O.B.E., M.A., D.Sc. . . . . . G. W. C. K.
Radiology.

LAMB, H., D.Sc., F.R.S. . . . . H. L.
Strings, Vibrations of.

LOWRY, THOMAS MARTIN, C.B.E., M.A., F.R.S. . . . T. M. L.
Magnetic Rotatory Power.
Quartz, Optical Rotatory Power of.
Tartaric Acid Rotatory Power.

MEES, C. E. KENNETH, D.Sc. . . . . . . C. E. K. M.
Light Filters.

MEGGERS, WILLIAM F., Ph.D., Physical Bureau of Standards, U.S.A. . . W. F. M.
Wave-lengths, Measurement of.

MERTON, T. R., M.A., D.Sc., F.R.S. . . . . . . T. R. M.
Spectroscopy, Modern.

OWEN, E. A., M.A., D.Sc., F.Inst.P. . . . . . . E. A. O.
Luminous Compounds.
Radioactivity.
Radium.

POWELL, HARRY J. (The late), C.B.E. . . . . . . H. J. P.
Glasses, Coloured.

RAYLEIGH, ROBERT JOHN, LORD, M.A., Sc.D., F.R.S. . . . . R.
Scattering of Light by Gases.

RHEINBERG, JULIUS, F.R.M.S., F.Inst.P. . . . . . J. R.
Glass, Platinised.
Graticules.

SAMPSON, R. A., D.Sc., F.R.S. . . . . . . R. A. S.
Telescope, The.

SCHOFIELD, F. H., B.A., B.Sc., ... F. H. S.
Tungsten Arc Lamp.

SHACKLETON, WILLIAM (The late), A.R.C.S., F.Inst.P. Revised by Professor Fowler, F.R.S. . . . . W. S.
Silvered Mirrors and Silvering.

SMITH, T., M.A., F.Inst.P. . . . . T. S.
Eyepieces.
Optical Calculations.
Optical Glass (with J. S. Anderson).
Periscopes.
Photographic Lenses.
Small Telescopes.
Transparent Substances (with J. S. Anderson).

SUTCLIFFE, JOHN H., O.B.E., F.Inst.P. . . . . . J. H. S.
Ophthalmic Optical Apparatus.

TUTTON, ALFRED EDWIN HOWARD, D.Sc., M.A., F.R.S. . . . A. E. H. T.

Crystallography.

TWYMAN, FRANK, F.Inst.P., and . . . . F. T.

GOLDSMITH, J. N., Ph.D., M.Sc., F.I.C. . . . . J. N. G.

Optical Rotation and the Polarimeter.

WALSH, JOHN WILLIAM TUDOR, M.A., M.Sc., F.Inst.P., A.M.I.E.E. . J. W. T. W.

Height-finder, The Paterson-Walsh.

Photometry and Illumination.

WILLIAMS, W. EWART, B.Sc., and . . . . . W. E. W.

TWYMAN, FRANK, F.Inst.P. . . . . . . F. T.

Polarimetry.

Saccharimetry.

WITHEY, WILLIAM HENRY, B.A. . . . . . W. H. W.

Glass, Chemical Decomposition of.

DICTIONARY OF APPLIED PHYSICS

— A —

Abbé Autocollimating Eyepiece. See "Goniometry," § (1).

Abbé Refractometer: an instrument for the rapid determination of refractive indices, principally of liquids. See "Spectroscopes and Refractometers," § (14).

Abbé Spherometer. See "Spherometry," § (3).

Abbé Theory of Microscopic Vision. See "Microscope, Optics of the," § (23).

Abbé's Autocollimating Spectroscope: a spectroscope for the determination of refractive indices of solids requiring the use of a minimum of material. See "Spectroscopes and Refractometers," § (12).

Abbé's Crystal Refractometer: an instrument primarily intended for the measurement of refractive indices of crystals, but suitable also for other substances. See "Spectroscopes and Refractometers," § (15).

Abbé's Method of adjusting Rangefinders. See "Rangefinder, Short-base," § (14).

Aberration, Spherical, stated by Lord Rayleigh to be unimportant when it nowhere exceeds a small fraction of the wavelength. See "Optical Parts, The Working of," § (3).
  General Theory of, for Lenses. See "Optical Calculations," §§ (8), (21), (22); also "Lens Systems, Aberrations of," § (5) (iv.).

Aberration of Camera Lenses, Spherical and Chromatic. See "Camera Lenses, Testing of," §§ (2), (3), and (11).

Aberration of Lenses. See "Optical Calculations," § (8); also "Telescope," § (3); "Lens Systems, Aberrations of," § (5); also "Photographic Lenses," § (8).

Aberrations, Measurement of. See "Objectives, Testing of Compound," § (3).
  Hartmann's method. See *ibid.* § (3) (iii.).
  Microscope method. See *ibid.* § (3) (i.).
  Shadow method. See *ibid.* § (3) (ii.).

Aberrations of Oblique Pencils in Microscope Systems. See "Microscope, Optics of the," § (19).

Aberrations of Telescopes, Practical Treatment of. See "Telescope," § (4).
  Von Seidel's Five Aberrations. See *ibid.* § (3).

Abrasion, Characteristic Features of. See "Optical Parts, The Working of," § (6).

Absorption Coefficients of X-rays. See "Radiology," § (17).

Accommodation: the process by which the eye adapts its focus to varying distances. See "Eye," § (20).

Achromatism of Telescope Lenses. See "Telescope," § (9).

Acoustics, Architectural: Jaeger's Treatment. See "Sound," § (22) (iii.).
  Marage's Work on halls in Paris. See *ibid.* § (22) (i.).
  Sabine's Work. See *ibid.* § (22) (i.).
  Watson's Investigations. See *ibid.* § (22) (iv.).

Active Deposit: a term used in radioactivity to denote the deposit of active matter on the surface of bodies which have been exposed for some time to the emanations from radio-elements. See "Radioactivity," § (18).

Acuity, Visual, the clearness with which the eye can see detail. See "Eye," § (28).

Adaptation, Dark: the highly sensitive condition of the eye when kept in darkness for some time. See "Eye," § (15).

Air, Dust-free, Experimental Observation of Scattering of Light by. See "Scattering of Light by Gases, etc.," § (3).

Aldis Spherometer. See "Spherometry," § (4).

— B —

BIPRISM, FRESNEL'S: an experimental arrangement for the production of interference fringes, using refraction through a biprism to form two adjacent virtual images of a source of light. See "Light, Interference of," § (4) (iv.).

BIQUARTZ: the name given to a device for measuring the rotation of the plane of polarisation of light by a substance; the device consists of two semicircular discs of quartz, 3·75 mm. thick, between parallel nicols, one half being right-handed quartz and the other left-handed quartz. See "Polarimetry," § (2).

BLACK BODIES AND RADIATION. See "Radiation Theory," § (4).

BLOCH'S METHOD OF CALCULATING AVERAGE ILLUMINATION. See "Photometry and Illumination," § (70).

BLONDEL'S PHOTOMETER: a photometer for the determination of average candle-power. See "Photometry and Illumination," § (43).

"BOILING" TUBE: an X-ray "gas" tube in which the anticathode is kept at a constant temperature by boiling water. See "Radiology," § (12).

BOLOMETER: an instrument for investigation of the infra-red spectrum; the temperature of a very fine strip of metal is measured in terms of the change of its electrical resistance, the strip being heated by radiation. See "Wave-lengths, The Measurement of," § (7).

BOMBARDON, BB♭ Monster: a brass wind instrument with valves. See "Sound," § (44).

E♭: a brass wind instrument with valves. See *ibid.* § (44).

BORIC ACID, renders glass resistant to water attack. See "Glass, Chemical Decomposition of," §§ (1) and (2).

BORIC OXIDE, USE OF, IN GLASS MANUFACTURE. See "Glass," § (11).

BOSANQUET'S CYCLE OF FIFTY-THREE: a particular musical temperament which gives as near an approach to just intonation as could be demanded from any instrument with fixed keys. See "Sound," § (6) (i.).

BOUGIE DÉCIMALE: a French standard of light. See "Photometry and Illumination," § (14).

BRACE SPECTROPHOTOMETER. See "Spectrophotometry," § (12).

BRACE'S POLARIMETER. See "Polarimetry," § (11) (iii.).

BRASHEAR'S OR SUGAR METHOD OF SILVERING MIRRORS; details of solutions and process. See "Silvered Mirrors and Silvering," § (2) (ii.).

BRASS BAND, VALVED INSTRUMENTS OF. See "Sound," § (44).

BREWSTER'S LAW. If light be incident upon the surface of a transparent body in such a direction that the tangent of the angle of incidence is equal to the refractive index of the body, then the reflected ray is plane polarised. The angle of incidence is in this case called the polarising angle. See "Polarised Light and its Applications," § (2).

BRIGHTNESS: used to denote the total visual effect of a light independently of its colour or purity. See "Eye," § (8).

Measured by the amount of light emitted by a surface per unit of projected area. See "Photometry and Illumination," § (2).

BROWNIAN MOVEMENT, use of ultramicroscope in study of. See "Ultramicroscope and its Application," § (2).

BUBBLES AS USED IN SPIRIT LEVELS. See "Spirit Levels," §§ (2), (3).

BUNSEN (GREASE-SPOT) PHOTOMETER: one of the early forms of photometer head still in current use. See "Photometry and Illumination," § (15).

β-RAYS, General Properties of. See "Radioactivity," § (13) (i.).

Magnetic spectrum of. See *ibid.* § (13) (ii.).

Scattering of, by matter. See *ibid.* § (13) (iv.).

— C —

CALIBRATION OF RANGE-SCALES OF RANGEFINDERS. See "Rangefinder, Short-base," § (15).

CAMERA, AERIAL. See "Photographic Apparatus," § (12).

Copying: a type of camera used for the photography of flat surfaces. See *ibid.* § (3).

Field or Portable. See *ibid.* § (2) (i.).

Hand: a general term for cameras devised especially for exposures of brief duration. See *ibid.* § (11).

Panoram: a type of camera used to photograph a large angular "stretch" of country. See *ibid.* § (7).

Photographic: in essential, a chamber for the purpose of exposing sensitive plates to the action of luminous images formed by optical means. The three main classes of cameras are (1) Stand Cameras, (2) Hand Cameras, (3) Aerial Cameras, and these again are classified into subgroups. See *ibid.* § (2).

Studio. See *ibid.* § (2) (ii.).

CAMERA LENSES, THE TESTING OF

The general problem of the testing of compound objectives is dealt with elsewhere.[1] The optical data that have to be determined in the case of photographic objectives are, generally speaking, more numerous than in the case of telescope objectives, and special methods of testing camera lenses have therefore been devised.

An idea of the general performance of a camera lens may be obtained by setting up an object, consisting of a network of straight lines on a white background,[2] and taking a photograph of it. The plate, after development, may then be examined by means of a microscope. In this way any distortion of the image may be detected and measured. The method, however, only gives a qualitative idea of any want of focus that may be present at any part of the plate.

A more accurate method is to make use of a specially designed lens bench, such as the one manufactured by Messrs. R. & J. Beck, Ltd. The principal features of this bench are illustrated in *Fig.* 1, which is reproduced from a photograph taken from the side, looking down at one end at an angle of about 45°. The various appliances for holding the lens, etc., slide on a steel dovetail-shaped bar B_1 which is supported on two struts (one of which is shown at *St*) provided with levelling screws. The lens L is screwed into one of a series of adapters held in a collar *Co* which revolves on ball bearings in the framework F_1. A clamping device *Cl* is provided for fixing the collar in any desired position. The framework can be moved along the slide S_1 by means of a rack and pinion actuated by the head H_1. This slide S_1 can be rotated about a vertical axis, the amount of rotation being measured on a scale Sc_1 on the platform Pl_1, which can be clamped to the bar B_1 at any convenient position. The framework F_2 can be moved along a slide S_2 by means of a rack and pinion actuated by the head H_2, the slide being fixed to the platform of the framework F_1. A bar B_2 is fitted at one end to a small carriage C_1 which can slide along the groove *Gr* in the framework F_2. This bar is constrained to move along the axis of the system between adjustable rollers R attached to the platform Pl_2, which can be clamped to the bar B_1 at any desired position. The observing microscope M can be moved along the axis relative to the bracket Br_1 (which is fixed to the bar B_2) by means of a rack and pinion actuated by the head H_3, the motion being registered by a scale and vernier. A vertical metal plate P, having a circular hole concentric with the axis of the system, is fixed to the small carriage C_1. When the lens holder, together with the framework F_2, is swung about the vertical axis of rotation, the plate remains parallel to the groove *Gr* and therefore normal to the axis of the lens. The distance between the axis of rotation and the plate P is measured by means of a millimetre scale Sc_2, one end of which fits into a bracket Br_2, so that the zero of the scale is always opposite the axis of rotation. A

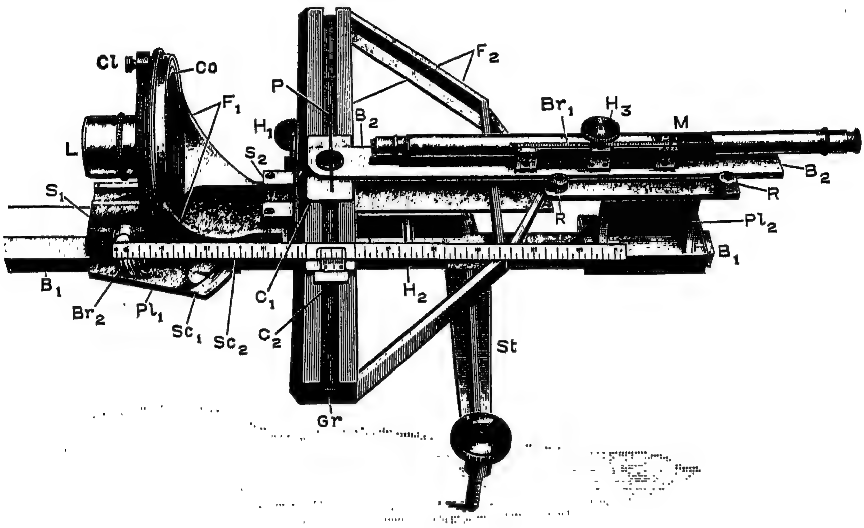


Fig. 1.

[1] See article on "Objectives, The Testing of Compound."

[2] See *Photographic Lenses*, by Conrad Beck and Herbert Andrews.

vernier is fixed to a small carriage C_2 which slides in the groove *Gr*, the vernier zero being in the plane of the left-hand surface of the plate P.

The optical data of a camera lens are determined as follows.

§ (1) Focal Length.—A piece of ground glass is held with its ground surface against the left-hand surface of the plate P and the observing microscope M is focussed on the ground surface. The framework F_2, carrying the microscope, is then moved along the slide S_2 by turning the head H_2 until the image of a distant object[1] is brought into the focal plane of the microscope.[2] The lens holder, together with the framework F_2, is next moved along the slide S_1 by turning the head H_1 until there is no lateral displacement of the image when the system is rotated through a small angle about the vertical axis. When this is the case, the vertical axis passes through a point, called the nul point, which divides the distance between the nodal points into parts whose ratio is equal to the lateral magnification. If the incident light is parallel, the nul point coincides with the second nodal point; the focal length is then given by the distance between the vertical axis and the focal plane, its value being read off on the scale Sc_2. When one is dealing with an object at a finite distance, the nul point is the point at which a line joining non-axial conjugate points in the object and image planes cuts the axis of the lens. In such a case it is advisable to make a separate determination of the position of the second nodal point by using parallel incident light. The back focal length is given by the distance between the last surface of the lens and the plate P.

§ (2) Spherical Aberration.—This is measured by placing a series of concentric stops in front of the aperture of the lens and measuring the movement of the microscope which is necessary to bring the image into focus for different zones of the lens. It is advisable to insert a colour filter between the microscope eyepiece and the observer's eye in order to eliminate the effects of chromatic aberration.

§ (3) Chromatic Aberration.—A series of colour filters, giving monochromatic regions more or less evenly spaced throughout the spectrum, is placed in turn behind the eyepiece of the microscope and the movements of the microscope necessary for bringing the different coloured images into focus are measured. It is sometimes advisable to stop out certain zones of the lens in order to eliminate the effects of spherical aberration.

§ (4) Coma.—A collimator, with an illuminated pinhole at its focus, is mounted on the axis of the system and the image of the pinhole is examined at different angles of rotation about the vertical axis. The dimensions of the comatic figure may be measured by means of a scale in the focal plane of the microscope eyepiece.

§ (5) Flare Spot.—The same arrangement is used as in the previous case, and the positions of flare-spot images are noted; such images do not give trouble unless they are in the neighbourhood of the focal plane of the lens.

§ (6) Astigmatism and Curvature of Field.—The system is rotated about the vertical axis, on which the second nodal point lies, through angles of 5°, 10°, etc., up to the maximum semi-angle that the lens is designed to cover, and at each angle (on either side of the axis) the positions of the microscope are determined for best focus of horizontal and vertical lines. The differences between these positions and the normal position of focus, that is, the focus for the centre of the field, give the distances of the positions of astigmatic foci from the plate (measured along the rays) at different parts of the field. The mean values of the distances of the foci for horizontal and vertical lines give a measure of the curvature of the field.

§ (7) Distortion.—The system is rotated as in the previous case, and the horizontal displacement of the image of a vertical line is measured for each angle by means of a horizontal graduated scale in the focal plane of the microscope eyepiece, the microscope being kept in the normal position. The horizontal displacement divided by the magnification of the microscope then gives the distortion in the plane of the plate.

§ (8) Effective Apertures of Stops.—The microscope is removed and a small pinhole, mounted in the plane of the plate P, is illuminated from behind. A piece of ground glass is held against the front of the lens mount and the diameters of the illuminated areas, corresponding to the different stops, are measured.

§ (9) Illumination at the Corners of the Plate.—The same arrangement is adopted as in the previous test, and the boundary of the illuminated portion is traced on the ground glass when the system is rotated through an angle corresponding to the semi-diameter of the plate to be covered by the lens. The area of this portion divided by the area of the unobscured aperture and multiplied by 100 and by the fourth power[3] of the cosine of the

[1] Either a set of suitable cross-lines at the focus of a well-corrected collimator objective mounted on the axis of the system, or a series of horizontal and vertical lines (black on white background and white on black background) at some considerable distance from the apparatus, may be used.

[2] On rotating the system about the vertical axis the left-hand surface of P moves in a plane which is equivalent to the plate of the camera.

[3] If the vignette is measured in a plane normal to the axis of the bench, the cube of the cosine is to be used.

angle then gives the percentage illumination at the corners of the plate relative to that at the centre; the loss of light due to oblique reflections and absorption further reduces the percentage illumination at the corners.

In addition to these measurements the relative centering of the components of the lens may be tested by using a small retinoscope for throwing a beam of light along the lens axis and examining the alignment of the images of the retinoscope hole, through which the observer looks. The number of glass-air surfaces in the lens may be determined by counting the number of these images. The presence of visible defects, such as striae, veins, etc., in the glass of the lens may be detected by mounting an illuminated pinhole at some distance from the lens and placing one's eye at the conjugate point. The aperture of the lens will then appear illuminated and any defects that may be present will be noticeable, especially if the image of the pinhole is partially obscured by a diaphragm.

§ (10) Definition.—It is customary to determine what is the largest stop for which the definition is satisfactory over the entire plate. In estimating the definition of a given lens it is advisable to compare with it the definition of a standard lens of similar dimensions.

§ (11) Photographic Methods.—The aberrations of a camera lens may be measured by a simple photographic method, which is employed at the Reichsanstalt.[1] The principle of the method is illustrated diagrammatically in *Fig.* 2. A slit S, illuminated by a bright source, such as a mercury vapour lamp, is placed at the focus of a collimator objective O. The camera lens L forms an image of this slit at S′, and a photographic plate P is mounted at S′ in such a position that it makes a small angle with the axis of L, the plane containing the slit and the optical axis being normal to the plate. The plate can be moved in its own plane at right angles to the axis of L, so that a series of exposures can be made for different zones of the lens and different wave-lengths. The various positions of foci are then determined by finding on the developed plate the places where the bundles of rays have the minimum breadths. In this way the spherical and chromatic aberrations of the lens may be determined. If the lens and the plate are rotated through different angles about an axis passing through the second nodal point of the lens, and a series of exposures made, the astigmatism, curvature of field, and distortion may similarly be determined. An estimate of the amount of coma present may be obtained from the degree of sharpness of the outer edge of the bundle of rays at the position of minimum breadth.

Another photographic method, which is applicable to the case of camera lenses, is the Hartmann method, particulars of which are given in the article "Objectives, The Testing of Compound."

The methods that have been described are not applicable to the case of process lenses, since these are specially designed for near work. In order to test such a lens a suitable object is mounted near the lens in such a way that it can be moved into different positions in a plane at right angles to the axis of the bench. The image is then examined with a microscope which can be moved parallel and at right angles to the bench axis and also about a vertical axis.

J. S. A.

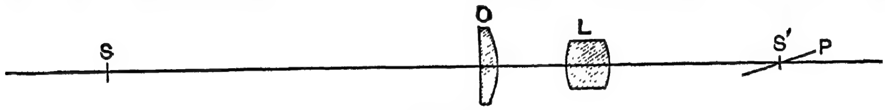


Fig. 2.

[1] *Zeits. Instrumentenk.*, 1915, xxxv. 104; 1916, xxxvi. 90; 1920, xl. 97.

Camera Shutters, Testing of. See "Shutters, Testing of Photographic."

Campbell and Smith's Method of testing Shutter Speeds. See "Shutters, Testing of Photographic."

Candle-power, the unit of luminous power. See "Photometry and Illumination," § (2). Of a Source of Light. See *ibid.* § (2).

Candle-power Variation with Voltage and Current in Electric Lamps. See "Photometry and Illumination," § (23).

Carbon Arc with specially treated Carbons, made to give an intrinsic brightness of 200,000 candles per square inch. See "Projection Apparatus," § (8).

Carcel Lamp, a flame standard officially adopted in France. See "Photometry and Illumination," § (8).

Carnotite, occurrence of, as ore of radium. See "Radium," § (2).

Cascade Method of Heterochromatic Photometry. See "Photometry and Illumination," § (104).

Cathode: the negative electrode in a discharge tube or an electrolytic cell. See "Radiology," § (9).

Cathode Rays: a stream of negatively charged electrons emitted with high velocity from the cathode when an

Coma: the name given to one of the aberrations from which the performance of an optical instrument may suffer. See "Telescope," § (3); "Optical Calculations," §§ (21), (22); "Lens Systems, Aberrations of," § (5) (iii.).

Measurement of. See "Camera Lenses, Testing of," § (4).

Compass Errors, method of determination. See "Navigation and Navigational Instruments," § (11).

Complete Flash: a term used to denote that a projector has the highest possible efficiency, *i.e.* that the whole front aperture of the projector, as seen from a distant point, is so filled with light as to possess the "intrinsic brightness" of the source. See "Projection Apparatus," § (3).

Concave Diffraction Grating: a form of optical grating invented by Rowland in 1881 to produce a focus spectra without the use of a lens and so to avoid chromatic difficulties in focussing and limitations due to absorption. See "Wave-lengths, The Measurement of," § (2).

Concave Grating Mounting, type introduced by Runge and Paschen. See "Wave-lengths, The Measurement of," § (2).

Condenser Lens: a lens used in projection apparatus to bend the directions of ray paths rather than to produce images. See "Projection Apparatus," § (13).

Conductivity, Electrical, of Glass. See "Glass," § (30).

Thermal, of glass, ·0015-·0025. See *ibid.* § (29).

Conical Refraction, External and Internal. See "Polarised Light and its Applications," § (7) (iii.).

Conjugate Points: the name given to a point and its geometrical image. See "Lenses, Theory of Simple," § (2), etc.

Their use in focal length measurements. See "Objectives, Testing of Compound," § (2) (ii.).

Constant Deviation Spectroscope. See "Spectroscopes and Refractometers," § (20).

Constant Potential, of a generator of X-rays. See "Radiology," § (27).

Constituents of Glass, their resistance to chemical attack. See "Glass, Chemical Decomposition of," § (2).

Contour (of optical surfaces). See "Interferometers, Technical Applications," § (4).

Contouring. See "Surveying and Surveying Instruments," § (37).

Contrast, Visual. See "Eye," § (16).

Cooke, Thomas: details of lenses designed by. See "Telescope," § (5).

Cooke Rangefinder. See "Rangefinder, Short-base," § (6).

Coolidge Tube, a hot-cathode X-ray tube. See "Radiology," § (13).

"Cords," a Defect in Glass. See "Striae." See also "Glass," § (16) (ii.).

Cornet, B♭: a brass wind instrument with valves. See "Sound," § (44).

Crova Method of Heterochromatic Photometry. See "Photometry and Illumination," § (103).

Crova's Spectrophotometer. See "Spectrophotometry," § (12).

Crystal Elements and Constructional Axes. See "Crystallography," § (4).

Crystal Faces, Miller's Method of distinguishing: a method of defining the relative positions of crystal faces in which the faces of a crystal are treated as a series of planes and any three of these, no two of which are parallel, are taken as planes of reference to which the positions of the other faces may be referred. See "Crystallography," § (3).

Crystal Structure, modern work on, based on the fact that the structure of a crystal is fundamentally that of a space-lattice, a three-dimensional trellis-work. See "Crystallography," § (11).

Analysis of, by X-rays: Laue's method, depending on the reflection and diffraction of X-rays by the molecules of a crystal, which are arranged on a space-lattice. See *ibid.* § (14).

Crystalline Lens: the lens of the eye. See "Eye," § (2).

CRYSTALLOGRAPHY

§ (1) Introduction.—The study of crystals can never be a matter of indifference to the physicist, for Crystallography is essentially the Physics of Solids. The higher branches of Optics deal very largely with crystals, and it is a crystal, tourmaline or calcite, that affords the best means of producing plane polarised light. Rock crystal, quartz, is the material of the best lenses employed by the optical investigator; a train of lenses and prisms of rock salt crystals are essential in the study of radiant heat; and the phenomenal progress which has recently been made in our knowledge of X-rays and of the ultimate structure of matter—the special domain of the physicist—is being very materially assisted by the use of crystals, the most perfectly organised form of solid matter. Yet the science of Crystallography has in the past been as strangely neglected by physicists as by chemists, to

whom the subject is likewise proving of prime importance, and has been left in the hands of a very few specialists, and regarded as a side issue chiefly connected with Mineralogy, oppressed by a forbidding kind of Mathematics and by a weird nomenclature. It will be the task of this brief summarised account of the subject to dispel this illusion, and to show how fascinating and intrinsically valuable it really is to the physicist.

Crystallography is essentially practical when its study is undertaken in a sensible manner. A few hours with a good crystal on a reflecting goniometer will bring out the chief properties of crystals and the laws of Crystallography so that they can never be forgotten, and will render the subject at once plain and simple. The goniometer, of which the best form is shown in *Fig.* 1, is nothing but the physicist's spectrometer, but with a small crystal, preferably no larger than a very small pea, as the object of study at the centre instead of the usual spectroscopic prism, and with a more delicate and convenient adjusting and centring apparatus. Also, instead of an ordinary parallel-jawed slit one is used which is expanded at its two ends, so that its image reflected from a crystal face exhibits a fine central part for accurate adjustment to the vertical cross spider-line, and broader ends which are brilliantly illuminated. It is surprising how brilliant such slit images are as seen in the telescope, when reflected from even the minutest faces, almost points, if the crystals are normally well developed.

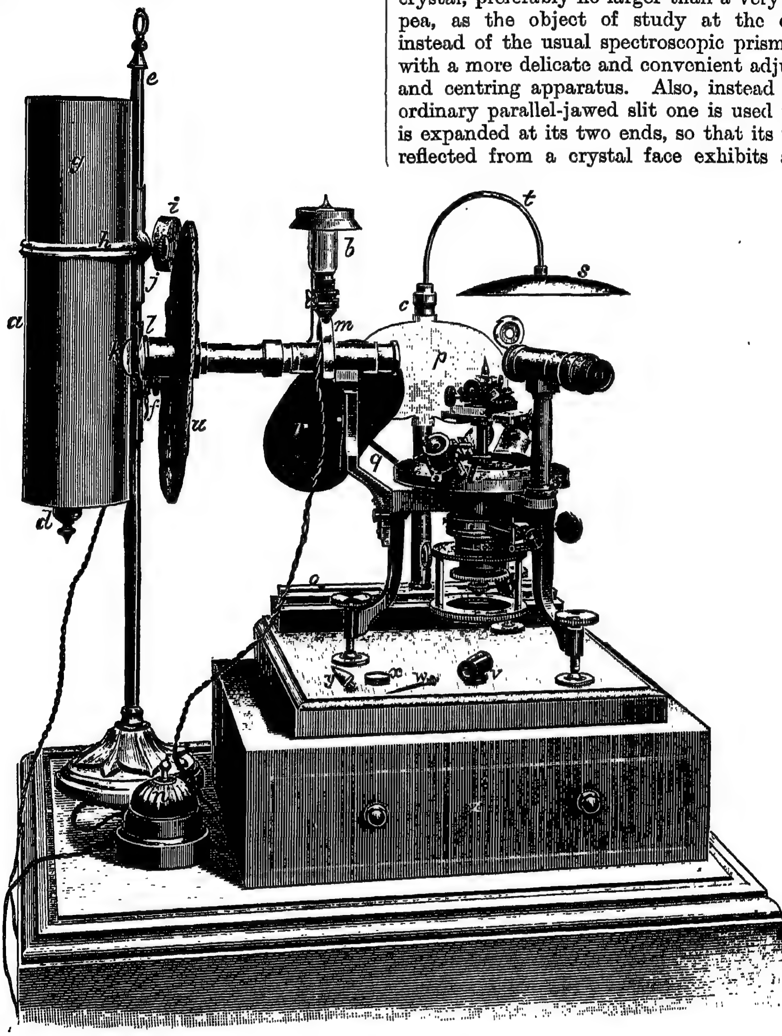


FIG. 1.—The Reflecting Goniometer in Use.

§ (2) NATURE OF CRYSTALS.—The first facts obviously and unmistakably impressed by such a practical investigation of a series of

well-developed crystals, of the same and of different substances, on the goniometer, are the following:

(*a*) Crystals are naturally formed solid polyhedra bounded by plane faces meeting in straight edges.

(*b*) These faces on the best crystals are true planes, highly polished and affording brilliant and sharp reflections of the goniometer signal-slit.

(*c*) The faces are inclined to each other at angles which usually display among themselves more or less symmetry, and are constant in value for the same substance; these particular values are thus characteristic of the substance, the only exception being when the symmetry is at its maximum, in the cubic system, when this perfect symmetry itself fixes the angles.

(*d*) The faces are arranged in zones, each zone being composed of faces parallel to a common axis, to which all their edges of intersection are also parallel. When this axis is set parallel to the axis of the goniometer, all the faces of the zone are automatically adjusted, so as to reflect the signal-image from every face in turn to the cross-wires of the telescope when the goniometer axis and crystal are rotated.

(*e*) When the signal-image from each successive face is adjusted to the cross-wires and the circle reading taken, the differences between the readings are the angles between the face-normals, and these are most conveniently considered as the interfacial angles; they are arranged in accordance with the symmetry developed.

§ (3) Miller's Method of distinguishing Crystal Faces.—A method of defining the relative positions of crystal faces by means of a compact symbol now universally adopted is due to W. H. Miller, Professor of Mineralogy at Cambridge from 1832 to 1881.

The faces of a crystal are a series of planes, and any three of these, no two of which are parallel, may be taken as planes of reference to which the position of the other faces may be referred.

(i.) *Axes of Reference.*—Consider then three planes passing through a point O which we take as origin and drawn parallel to any three faces of the crystal; these planes intersect in three straight lines OX, OY, OZ (*Fig.* 2); these lines we take as axes of reference.

The planes may be parallel to any three faces, but it will usually be possible to find three which, from the regularity of the crystal with regard to them, are clearly of importance in its construction, and in a number of cases two or more planes can be found mutually at right angles.

Any fourth plane will cut the axes in points A, B, C, and its position can be defined by the lengths of the intercepts OA, OB, and OC—or rather, since we are not concerned with the actual location of the face, which depends on the size and not merely the form of the crystal—by the ratios OA : OB : OC of the three intercepts, which thus fix a series of parallel planes. It is now a universal (international) convention for the description of crystals that the axis OY (and its intercept OB) shall run from left to right, and the OX axis (OA intercept) from front to back.

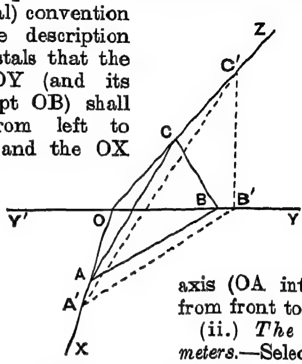


Fig. 2.

(ii.) *The Parameters.*—Select then any such plane and let the lengths of the intercepts be a, b, c respectively. We define this as a parametral plane, and the values of a, b, c measured in any convenient unit as the parameters. It will usually be found that there is a plane, developed as a prominent face, for which some simple relations exist between a, b, and c, *e.g.* all three parameters are equal, or two of them are equal and differ from the third; it is convenient but not essential to select such a plane as the parametral plane.

(iii.) *The Indices.*—A plane parallel to any other face of the crystal will cut the axes in three points, A′, B′, C′ say.

Then it is found by observation that the intercepts OA′, OB′, OC′ are proportional respectively to a/h, b/k, c/l, where h, k, l are in all cases small whole numbers—including zero—in few cases so great as 6. These numbers are called the indices of the plane, and if we know in any case the position of the axes and the values of the parameters the position of any other plane is defined by its indices and is denoted by the symbol $(h\,k\,l)$.

This implies that if the known parameters be a, b, c, and we cut off on the axes lengths proportional to a/h, b/k, and c/l, the face in question will be parallel to the plane so defined.

The relative positions of the axes are fixed by the angles between them. If we put $YOZ=\alpha$, $ZOX=\beta$, $XOY=\gamma$, we see that the position of any face is determined by the values of the three parameters a, b, c, the three angles α, β, γ, and the three indices h, k, l; moreover, in all cases h, k, l are small integers including possibly zero.

A face may of course cut the axes on either the positive or the negative side of the origin, *i.e.* either between O and X or on the side of O remote from X. This will be indicated by giving a negative sign to the corresponding index, which becomes $-h$ or, as it is usually written, $\bar{h}$. Again, if one of the indices, say h, is zero, the corresponding intercept is $a/0$ or infinity. This indicates that the axis OX is parallel to the face in question.

Again, since we are only concerned with the ratios $a:b:c$ we may take any one of these quantities to be unity and refer the others to it; it is usual to select b as unity, so that the ratio becomes $a:1:c$.

§ (4) CRYSTAL ELEMENTS—CONSTRUCTIONAL AXES. — These axes are spoken of as the constructional axes of the crystal, and the various systems of crystals are grouped according to the values of the parameters a, b, c, expressed as $a:1:c$, and the angles α, β, γ, between the constructional axes. These quantities are known as the crystal elements.

§ (5) CRYSTAL SYSTEMS.—These are seven in number. Each system contains a number of classes characterised by a common set of constructional axes, *i.e.* common type of crystal elements, and certain common features of symmetry.

§ (6) ELEMENTS OF SYMMETRY. — The elements of symmetry which belong to a crystal are two in number; it may possess a plane or planes of symmetry or it may be characterised by an axis or axes of symmetry, or by both elements.

(i.) *Plane of Symmetry.*—Imagine a crystal to be divided into two portions, A and B, by a plane which we suppose capable of producing a reflected image of A; let B′ be this image. In general there will be no resemblance between B and B′, but if it should happen that each face of B′ coincides with or is parallel to a face of B, then the crystal has been divided into two symmetrical portions by the plane, the plane is a plane of symmetry.

(ii.) *Axis of Symmetry.*—Imagine now a line in a crystal possessing the property that if the crystal be rotated about this line through some definite angle each face is brought into a position either coincident with or parallel to that occupied by some other face before the rotation took place. The crystal is said to be symmetrical about the line, and the line is an axis of symmetry. If the crystal were mounted on a goniometer with the axis of symmetry coincident with the axis of the goniometer it would not be possible to infer from the readings of the instrument that the crystal had been moved. Thus, for example, in a cube each of the faces of the cube is a plane of symmetry and each of the edges an axis of symmetry; a cube, however, has many more planes and axes of symmetry than these. Or consider a crystal bounded by two regular figures, ABCDE, A′BCDE (*Fig.* 3), having the rectangular face BCDE in common, and the eight triangular faces equal; the face BCDE is clearly a plane of symmetry, and the line AA′ which is perpendicular to it is an axis of symmetry.

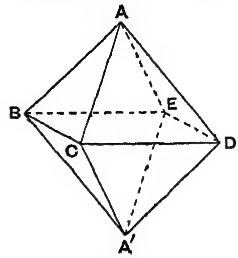


FIG. 3.

The angle through which rotation takes place in order to bring the crystal into a symmetrical position is in all cases some submultiple of 360°, corresponding to division of the whole circle by two, three, four, or six. If the rotation be through 180°—360°/2—the symmetry is *digonal*, for 120°—360°/3—it is *trigonal*, for 90°—360°/4—it is *tetragonal*, and for 60°—360°/6—the symmetry is *hexagonal.*

In some crystals a more complicated type of symmetry exists in which we can bring the crystal into the symmetrical position by supposing (1) that it is rotated about an axis, and (2) that each face is displaced as it would be if reflected in a plane; we have to consider the combined effect of an axis of symmetry and a plane of symmetry.

Thus consider a crystal of which a section is given by ABCD (*Fig.* 4), a quadrilateral figure in which

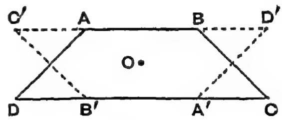


FIG. 4.

AB and CD are parallel and BC and AD equally inclined to AB. On rotating this through 180° about an axis through O, the mid point of the bisector of the two parallel sides, perpendicular to the paper, we get A′B′C′D′, which has not all its faces parallel to those of the original crystal, but by supposing B′C′ reflected in a plane through the middle point of AD perpendicular to the paper, and A′D′ similarly reflected in a plane through the mid point of BC, we recover the original crystal.

§ (7) THE SEVEN SYSTEMS OF CRYSTALS.—Before discussing the types of symmetry which characterise the various systems of crystals each of which has its own type of crystal elements defined by the parameters a, b, c, and the angles α, β, γ, of the constructional axes, it will be useful to give in a schedule the relations between these quan-

tities for each of the seven systems, which are named as shown in Table I.

TABLE I

SYSTEMS OF CRYSTALS

Name.	Parameters.	Interaxial Angles.
i. Triclinic . .	$a \gtrless b \gtrless c$	$\alpha \gtrless \beta \gtrless \gamma$
ii. Monoclinic .	$a \gtrless b \gtrless c$	$\alpha = \gamma = 90°$ $\beta \gtrless 90°$
iii. Rhombic . .	$a \gtrless b \gtrless c$	$\alpha = \beta = \gamma = 90°$
iv. Tetragonal .	$a = b \gtrless c$	$\alpha = \beta = \gamma = 90°$
v. Hexagonal .	$a = b \gtrless c$	$\alpha = \beta = 90°$ $\gamma = 60°$ and $120°$
vi. Trigonal . .	$a = b = c$	$\alpha = \beta = \gamma \gtrless 90°$
vii. Cubic . .	$a = b = c$	$\alpha = \beta = \gamma = 90°$

§ (8) CLASSES OF SYMMETRY. — Each of these seven systems includes crystals of various classes of symmetry, in all thirty-two in number, which are built up from the elements of symmetry—the plane of symmetry and the axis of symmetry—already described.

(i.) The *Triclinic System*, which is the least symmetric, comprises two classes. The three axes are unequal and unequally inclined. Class 1 has no symmetry whatsoever, every face being a "form" unto itself, that is, it has no fellow or fellows of like relation to the constructive axes about which the crystal can be imagined to be erected. Class 2, however, is symmetrical about the centre; thus the crystal is composed of parallel faces, each pair of which is a form in the sense just alluded to. One of the best-known triclinic substances is copper sulphate, $CuSO_4 . 5H_2O$, a crystal of which is shown in *Fig.* 5.

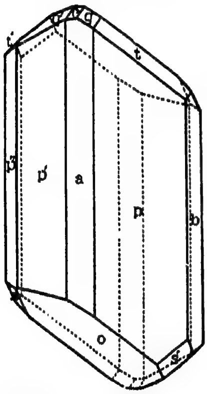


FIG. 5.—A Crystal of Copper-sulphate.

(ii.) The *Monoclinic* or *Monosymmetric System* comprises three classes. The three axes are unequal, two being inclined to each other at an angle other than 90°, while the third, always chosen as the *b* axis, is normal to their plane. Two real elements of symmetry are now possible, a plane of symmetry, that containing the inclined axes, and a digonal axis of symmetry perpendicular thereto and identical with the normal construction axis. Class 5 possesses both these elements of symmetry, while Class 3 is endowed with the plane of symmetry only, and Class 4 with the digonal axis only. An excellent example of full monoclinic Class 5 symmetry is potassium nickel sulphate, $K_2Ni(SO_4)_2 . 6H_2O$, a typical crystal of which is shown in *Fig.* 6.

(iii.) The *Rhombic* or *Orthorhombic System* includes three classes. The three axes are unequal but are arranged at angles of 90° to each other. One class, No. 8, possesses all the six elements of symmetry which now become possible, namely, three planes of symmetry, which are also the three constructional axial planes, and three digonal axes of symmetry, identical with the three rectangular axes and the intersections of the planes. Potassium sulphate, K_2SO_4, is a common substance crystallising in this Class 8, and a characteristic crystal is represented in

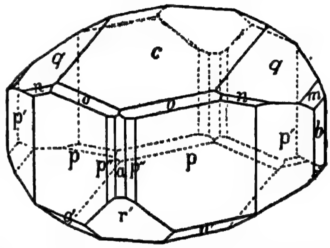


FIG. 6.—A Crystal of Potassium Nickel Sulphate.

Fig. 7. Class 6 possesses the three axes of symmetry only, and Class 7 possesses two of the planes and one axis of symmetry, the axis in which the planes intersect.

(iv.) The *Tetragonal System* goes a step further in degree of symmetry. Two of the axes are of equal length, the third being unequal; the angles are all equal to 90°. In describing or setting up the crystal the unequal axis is made vertical. No less than seven classes are now possible, all of which possess the systematic characteristic, a tetragonal axis of symmetry which is identical with the vertical constructional axis. Class 9 possesses this symmetry element alone, but Class 15, the highest of the system, is also endowed with four symmetry planes intersecting each other at 45° in the tetragonal vertical axis, an equatorial symmetry plane perpendicular to the vertical axis, and four digonal axes of symmetry lying in that

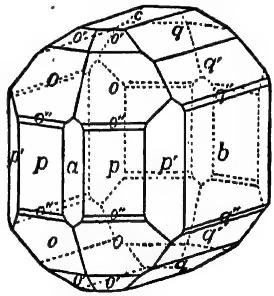


FIG. 7.—A Crystal of Potassium Sulphate.

equatorial plane. Zircon, silicate of zirconium, $ZrSiO_4$, crystallises with the symmetry of this class, and a typical zircon is portrayed in *Fig.* 8. Between this "holohedral" class (of full symmetry) and Class 9 are five other classes possessing various combinations of some of these symmetry elements but always including the tetragonal axis, the essential symmetry element of the tetragonal system.

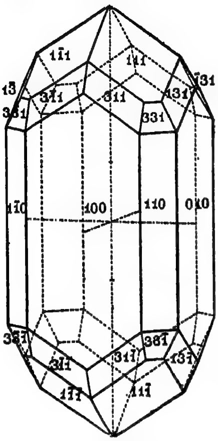


FIG. 8.—A Zircon Crystal.

(v.) The *Hexagonal System* is similar in many ways to the tetragonal, but the equal axes are inclined to each other at 60° instead of 90°; it is characterised by a hexagonal vertical axis of symmetry. There are five classes in the system, the simplest, Class 23, possessing only the essential hexagonal axis of symmetry; but the most symmetrical, the holohedral Class 27, has six planes of symmetry intersecting each other at 30° in the hexagonal axis, and an equatorial plane of symmetry with six digonal axes of symmetry lying in it at 30° from each other, a total of fourteen elements of symmetry. Beryl, $Be_3Al_2(SiO_3)_6$, the beautiful gemstone known as aquamarine when pale green in colour and emerald when dark green, is an excellent example of Class 27, and a typical beryl crystal in shown in *Fig.* 9.

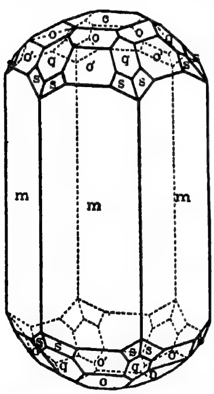


FIG. 9.—A Beryl Crystal.

(vi.) The *Trigonal System* has three equal and equally inclined axes, generally at other than 90°. The axes, indeed, are the edges of a rhombohedron, which resembles a cube compressed or extended along one diagonal, and this diagonal is the vertical axis of trigonal symmetry and not a construction axis; it is the essential element of symmetry of the system, which includes seven classes, the simplest, Class 16, having this as its only symmetry element. The most symmetrical, the holohedral Class 22, possesses in addition three symmetry planes intersecting at 60° in the trigonal axis, an equatorial symmetry plane, and three digonal axes lying therein at 60°. No well-known substance exhibits this full trigonal symmetry, although two rarer substances, including the gemstone benitoite, $BaTiSi_3O_9$, belong to the class. But two very important minerals, calcite, $CaCO_3$, and quartz, SiO_2, belong to Classes 21 and 18 respectively, and figures of typical specimens of calcite and quartz crystals are shown in *Figs.* 10, 11, and 12. *Fig.* 10 also shows the three constructional rhombohedral axes, and the vertical trigonal axis. This Class 21 only differs from the holohedral Class 22 by having no equatorial plane of symmetry. The actual crystal shown in *Fig.* 10 is a combination of the rhombohedron, with edges parallel to the constructional axes, and the hexagonal prism parallel to the vertical axis.

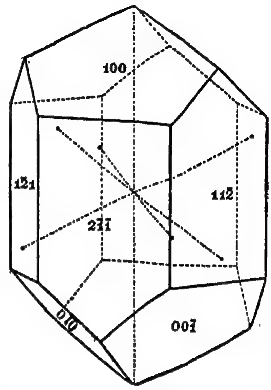


FIG. 10.—A Crystal of Calcite.

The Quartz Class 18 has all the four axes of symmetry (three horizontal and one vertical) of the system, but no symmetry planes. Now it is an interesting and highly important fact that all the classes of symmetry, and there are eleven of them, which possess no plane of symmetry are distinguished by the property of right- and left-handedness, or, as it is termed, "enantiomorphism." Two varieties of crystals, one the mirror image of the other, right- and left-handed like a pair of gloves, are not only possible but are often found developed, and two such crystals of quartz are portrayed in *Figs.* 11 (left-handed) and 12 (right-handed). Moreover, to enhance the interest, it is just this remarkable property which is possessed by the crystals of all substances which rotate

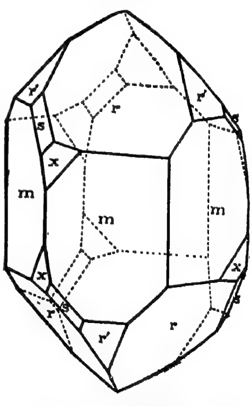


FIG. 11.—A Left-handed Crystal of Quartz.

the plane of polarisation of plane-polarised light, and if a section-plate were cut out of each of the two crystals shown, perpendicularly to the vertical axis, it would rotate the plane of polarisation to the left in case of *Fig.* 11 and to the right in the case of *Fig.* 12.

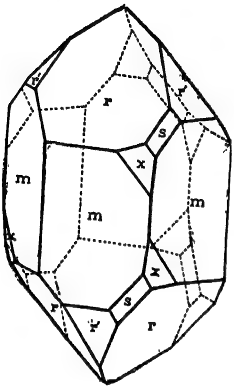


FIG. 12.—A Right-handed Crystal of Quartz.

(vii.) The *Cubic System* of highest symmetry is typified by the cube itself, the three constructional axes being parallel to the cube edges, equal in length and all mutually at right angles. There are five classes comprised in the system, the least symmetric, Class 28, possessing three rectangular digonal axes of symmetry coincident with the constructional axes, and four trigonal symmetry axes equally inclined to the former. But the most symmetric, the holohedral Class 32, possesses no less than twenty-two elements of symmetry, namely, the three already mentioned as digonal axes but which are now tetragonal ones, and the four trigonal axes also already referred to, and in addition six digonal axes bisecting the angles between the tetragonal axes, three planes of symmetry (the cube planes) perpendicular to the tetragonal axes, and six other symmetry planes bisecting the angles between the three just

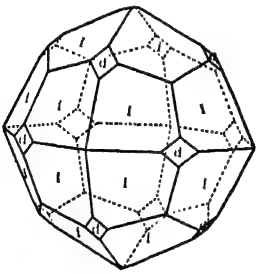


FIG. 13.—A Crystal of Garnet.

mentioned. The ancient Greeks may well have taken the cube as their symbol of perfection, their geometricians having understood the immense possibilities of symmetry incipient within it. Garnet, the beautiful yet very common silicate of the general composition $R''_3R'''_2(SiO_4)_3$, in which R'' stands for calcium, magnesium, ferrous iron, or manganese, and R''' for aluminium, ferric iron, or chromium, crystallises in this Class 32 of maximum symmetry, and a typical red garnet measured by the writer is shown in *Fig.* 13. The cubic system is unique in that all the faces,

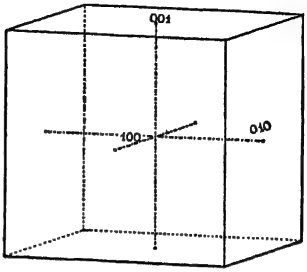


FIG. 14.—The Cube.

and they are exceedingly numerous, possible to be developed on the crystals conforming to the system are inclined at angles absolutely fixed by the symmetry and invariable. Its three

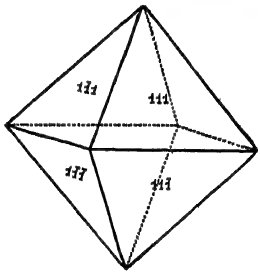


FIG. 15.—The Octahedron.

simplest forms, the cube, octahedron, and rhombic dodecahedron, represented in *Figs.* 14, 15, and 16, are also unique in that there can be only one cube, one octahedron, one rhombic

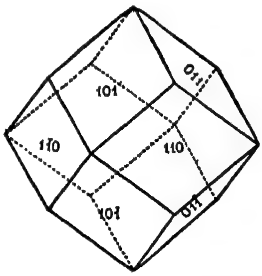


FIG. 16.—The Rhombic Dodecahedron.

dodecahedron. But there are four other holohedral cubic forms, the hexakis octahedron, the icositetrahedron, the triakis octahedron, and the tetrakis hexahedron, represented by

the next four figures, *Figs.* 17, 18, 19, and 20, which may have several representatives, the commonest of which are specified in (§) 10, yet each of which has its own definite angles fixed

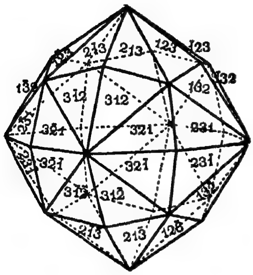


FIG. 17.—The Hexakis Octahedron.

by the symmetry. Also, there are ten other polyhedra characteristic of the cubic classes of lower than the holohedral symmetry, of which two may be taken as examples, the tetrahedron and the pentagonal dodecahedron,

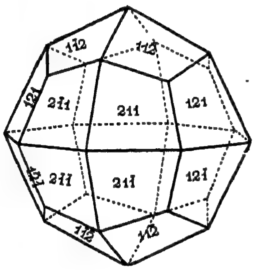


FIG. 18.—The Icositetrahedron.

shown in *Figs.* 21 and 22. The tetrahedron is a form of both Class 31 and Class 28, the former class being produced when the three axial planes are no longer planes of symmetry and the six digonal axes are no longer operative.

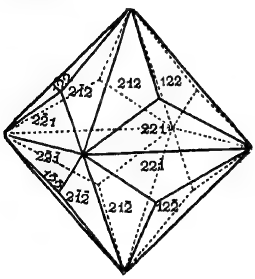


FIG. 19.—The Triakis Octahedron.

The pentagonal dodecahedron is most characteristic of Class 30 (in which iron pyrites, FeS_2, crystallises), which is formed when the six diagonal planes of symmetry are eliminated, the three cubic planes being operative; the form occurs, however, also in Class 28. The

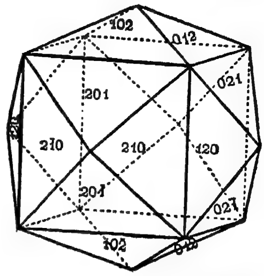


FIG. 20.—The Tetrakis Hexahedron.

polyhedron represented in *Fig.* 17, the hexakis octahedron, possesses the large number of 48 faces. It is the solid produced by the operation of the whole 22 symmetry elements when a single face inclined unequally to all three cubic axes is taken as the start.

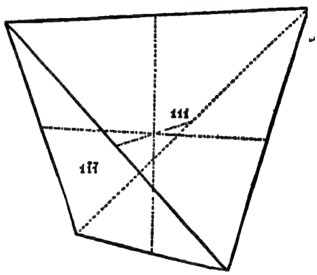


FIG. 21.—The Tetrahedron.

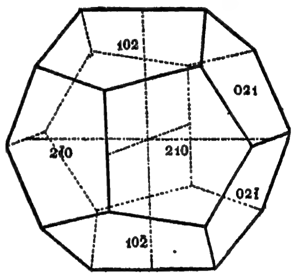


FIG. 22.—The Pentagonal Dodecahedron.

§ (9) THE REAL ELEMENTS OF SYMMETRY.—It has been essential to refer to these geometrical details concerning the seven systems of crystal architecture, and the 32 classes of symmetry combinations which they comprise, in order to render quite clear one of the greatest advances of the last third of a century. Previous to that the existence of classes of less than the full "holohedral" systematic symmetry had been accounted

for by assuming that the half or three-quarters of the full number of faces corresponding to the holohedral class were suppressed, or latent. For instance, the suppression of alternate faces of the octahedron was supposed to give the tetrahedron; and the elimination of half of the faces of the tetrakis hexahedron to give the pentagonal dodecahedron. These classes were therefore denominated "hemihedral" or "tetartohedral," according as one-half or only one-fourth of the full number of faces were present. But the work of Victor von Lang, von Groth, Story Maskelyne, and Viola effectually eliminated this feeble and very unscientific supposition, by showing that there are certain definite elements of symmetry, relatively few in number—those in fact which have been specified in § (6)—and that each crystal class of the 32 is endowed with its own invariable and fixed number and character of symmetry elements, and that all of them are active, the crystal being the outcome of the operation of all the symmetry elements specified as governing and forming the class.

§ (10) The Law of Rational Intercepts or Indices.—When we proceed to determine the indices of the other forms besides the primary, or parametral, and their individual faces on the crystal, we are saved from complicated figures and fractional or decimal values by a most remarkable law, already mentioned, which has proved to be the key to crystal structure. When we have made the comparison and got out our indices, they all prove to be simple integers. That is, the faces developed on crystals by Nature are only such as have rational, integral, intercepts, whole numbers only, and usually very low ones such as 2, 3, 4, 5, or possibly 6, and but very rarely higher numbers. In other words, we have not to deal with faces of every possible angle, but only with faces having angles arranged at definite intervals. Moreover, if symmetry be present, as is usual, the faces will form the natural groups or "forms" already referred to, for which the index numbers are the same, all these faces of any one form having the same relation to the symmetry developed. We have seen that in the system of highest symmetry, the cubic, for which the parameters are all equal, the general (faces inclined to all three axes) form, the hexakis octahedron, has no less than 48 faces, all represented by the same form symbol, {321} for instance. The kind of bracket just used is, in fact, reserved to enclose a "form" symbol, the indices given as representing the form being those of a face in the top right front octant. At the other extreme, in Class 1, possessing no symmetry, every face is a form unto itself.

The regular octahedron, of 8 faces as its name implies, is {111}. The rhombic dodecahedron is {110}, as each face is parallel to one cubic axis and cuts off equal intercepts from the other two axes; it possesses 12 faces as again implied by its name. The cube is {100}, for each face is parallel to two axes; (100) is the front face, (010) the right face, and (001) the top face. The ordinary form of simple bracket here used is that reserved for the symbol of a face. The other four forms represented in *Figs.* 17 to 20 are known to have several representatives, the commonest being, for the hexakis octahedron {321}, shown in *Fig.* 17, and {421}, all three indices being always different; for the icositetrahedron of 24 faces {211}, shown in *Fig.* 18, and {311}, two indices being always equal and less than the third; for the triakis octahedron also of 24 faces {221}, shown in *Fig.* 19, and {331}, the two equal indices being always greater than the third; and for the tetrakis hexahedron {210}, shown in *Fig.* 20, and {310}, every face of the 24 which it possesses being parallel to one axis and differently inclined to the other two.

In no other system than the cubic are the whole of the angles determined by the symmetry itself, and able to be calculated directly therefrom by spherical trigonometry. In the other systems the angles not fixed by the symmetry require to be determined by goniometrical measurement, from the results of which the crystal elements can be directly calculated. The occurrence of the faces in zones is a great help, especially as most faces lie at the intersection of two or more zones. Indeed the position of the faces can be located on a sphere directly from the measurements of the angles in the various zones; and by a very simple construction, the points where normals to the faces from the common centre of crystal and sphere cut the sphere can be projected on to a plane, the eye being supposed to be at the north or south pole and the plane of projection to be the equatorial plane. Such a projection on paper, the Stereographic Projection, affords us a concise and invaluable plan of the crystal, and the spherical triangles on it indicate to us the obvious course of the calculations, by which the crystal elements and the interfacial angles themselves can all be computed, provided one, two, three or five (according to the degree of symmetry) "basal" angles are measured as the basis of calculation.

Thus *Fig.* 23 represents the stereographic projection of a crystal of topaz $(AlF)_2SiO_4$, which is orthorhombic, of the holohedral Class 8. Every zone is represented by a circular arc or straight line, and in the case of the zone of faces parallel to the vertical axis, by the outer complete circle. The points on each zone represent the positions of the

various faces indicated by the Millerian indices. The projection is that of the upper hemisphere, but as the plane of the paper is a plane of symmetry of this rhombic crystal the poles on the lower hemisphere are identically placed. The two vertical planes of symmetry are represented by the two diameters parallel to the page edges. The symmetry of the crystal is thus clearly and fully indicated by the stereographic projection.

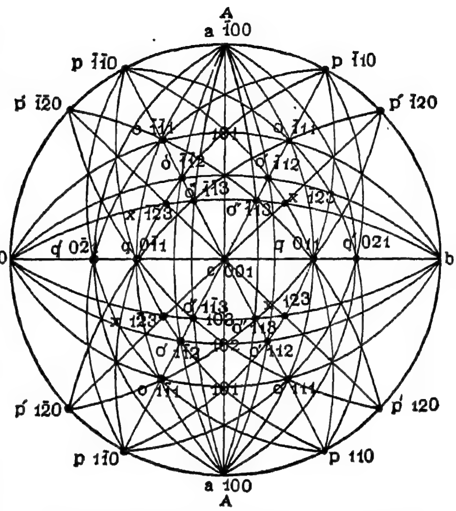


FIG. 23.—The Stereographic Projection of Topaz.

§ (11) CRYSTAL STRUCTURE.—Now the limitation of the number of possible forms to such as have rational indices is of prime importance with regard to the structure of crystals. For it means that the crystal is built up of units of definite appreciable size, the "bricks" of the crystal edifice. The Abbe Haüy, who first recognised the law at the time of the French Revolution, imagined them as "molécules intégrantes," and since his time the idea has developed, very slowly for many years, but lately much more rapidly, until a geometrical theory of crystal structure has been evolved which has probably now reached finality, having been confirmed in a remarkable manner by the new X-ray analysis of crystals presently to be referred to. Its essence is as follows:

Considering the elementary atoms composing the crystal substance as points, the crystal structure is that of a homogeneous arrangement of points. There are 230 modes of arranging points in a homogeneous structure, such as is possible to crystals, having regard to all the limitations which are imposed by the law of rationality and the fact that crystals have plane faces. Of these, 65 are point-systems of the simpler kind, involving geometrical "space operations" of only the first order (rotations about axes only). They were first described by L. Sohncke. The remaining 165 more complicated point-systems involve space operations of the second kind (reflections over planes). These were recognised and described simultaneously by E. S. Fedorov, A. Schoenflies, and W. Barlow. Fundamental to the whole 230, however, are 14 space-lattices defined by Frankenheim and Bravais, and many of the 65 Sohncke systems reduce to these space-lattices when groups of the points are considered as units, or are replaced each by a single point; in a few cases the space-lattices are special cases of the point-systems.

We thus come to the basal fact that the structure of a crystal is fundamentally that of a space-lattice, a three-dimensional trellis-work. Three of these space-lattices are of cubic symmetry, the points being arranged as the simple cube, the centred cube, and the face-centred cube. They are shown in *Figs.* 24, 25, and 26. Two others are tetragonal, four are rhombic, two monoclinic, one triclinic, one is hexagonal, or trigonal-prismatic, and another is rhombohedral-trigonal. In the crystals of very simple chemical compounds, and of the chemical elements themselves, the space-lattice is directly formed by the chemical atoms. In the more complicated crystalline substances, when more atoms go to the molecule, the space-lattice points are surrounded or replaced by groups of atoms. In any case, the structure is that of one of the 230 systems of points, according to which

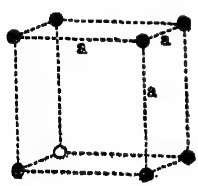


FIG. 24.—The Cubic Space-lattice.

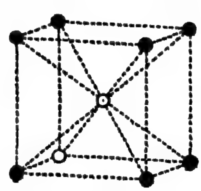


FIG. 25.—The Centred Cube Space-lattice.

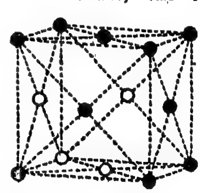


FIG. 26.—The Face-centred Cubic Space-lattice.

alone homogeneous crystal structures must be arranged, or is a special and very simple case of one of them. The crystal faces are parallel to the various planes of atoms of the space-lattice. Any three adjacent points of the lattice of course determine a face. It is, indeed, only these possible planes of atoms, the nodes of the space-lattice, corresponding to all the possible combinations of three adjacent atom-nodes, that are possible faces, and this is the very simple explanation of the law of rational indices or intercepts. In short, the parallelepipeda formed by joining the rows of points of the trellis by straight lines, along the three dimensions of space, the unit blocks of the three-dimensional lattice, are the "bricks" of the crystal edifice. Further, the Cleavage Planes of crystals are those along which the atomic points or nodes are most densely packed; for obviously in the perpendicular direction the cohesion must be less, the atomic attractive points being further apart, and the crystal is more readily torn asunder by a force thus perpendicularly directed. A very simple and rational explanation of the very important property of cleavage, which so many crystals exhibit, is thus afforded.

§ (12) Relative Measurement of Space-lattice Cells in similarly constructed Crystals.—The points at the corners of the unit parallelepipedon of the space-lattice may be either the atoms themselves (simple cases), or the representatives of groups of atoms; or we may regard them as representing the chemical molecules or small groups of molecules (not usually more than four, but occasionally as many as eight) which form the complete point-system of the structure, and thus we come to regard the parallelepipedon as the cell or habitat of such molecular or polymolecular structural unit. Now in a series of substances crystallising similarly—those known as "isomorphous"—in which the only chemical difference is that one chemical element is replaced by another belonging to the same family group of the periodic system, it must be obvious that the structure is a strictly analogous one, belonging to the same crystal-class and exhibiting the same "forms," and only differing in the dimensions of its parallelepipeda. It was suggested by F. Becke in the year 1893 that a relative measure of these cell dimensions might be obtained by combining the crystallographic axial ratios with determinations of the density of the crystals. For the latter divided into the molecular weight of the substance affords the Molecular Volume of Kopp, and this may be regarded as the volume of the cell. Combination of this with the axial relative lengths should therefore yield us the relative edge dimensions of the cell. The suggestion was taken up practically by Muthmann and by the writer simultaneously and independently in the year 1894, and applied by the former to the case of the rhombic permanganates of the alkalies, the crystals of which he had measured and submitted to density determinations, and by the latter to the rhombic sulphates and selenates of potassium, rubidium, caesium, and ammonium, as well as to a large number of the double sulphates and selenates of the well-known monoclinic series with six molecules of water of crystallisation, of which these alkali salts are the dominating constituents. The new Space-Ratios thus obtained were called "Topic Axial Ratios" by Muthmann (from τόπος, space), and "Distance Ratios" by the writer. Now the interesting result was obtained that these relative space-ratios indicated a regular increase in the space-lattice cell volume and edge dimensions, as the atoms of potassium were replaced by the heavier atoms of rubidium, and these in turn by the still heavier ones of caesium. The actual values will be found given in § (17) (iv.), for the simple rhombic sulphates of the alkalies. It was further indicated that the volume and edge dimensions of the cells of the analogous ammonium salt were almost exactly identical with those of the corresponding rubidium salt, the intermediate member of the group of salts. It has to be remembered that the atomic weight (84·9) and atomic number (37) of rubidium are practically exactly midway between these constants for potassium (38·9 and 19) and caesium (131·9 and 55). Hence, the crystals of these ammonium and rubidium salts are essentially isostructural.

Now these interesting results were in keeping with those derived from the writer's previous researches. For it had been shown that the interfacial angles and elements of the crystals of these isomorphous series exhibit a similar progression, corresponding to the advance in atomic weight and atomic number, and in a long series of subsequent memoirs (just completed, 1922) it has been proved conclusively that this is a general law for all the crystal properties, morphological, optical, and thermal, of these important rhombic and monoclinic series of isomorphous salts. The volumes and edge dimensions of their space-lattice cells thus conform to the general law of progression which the writer has now established for these series.

§ (13) Final Proof of Haüy's Law of Constancy of Crystal Angles.—Incidentally this result has also definitely settled the long-argued contradiction between the view of Haüy—that every chemical substance was characterised by its own crystalline form—and the principle of isomorphism of Mitscherlich, who in first putting forward his discovery of the principle in its cruder form

held that isomorphous substances are equiangular, and therefore that quite a large number of similarly constituted substances had not only the same form but also the same angles. But the results of the writer's work prove that there are small but very real differences in the crystal angles, rarely reaching 3° and often less than 1°, as well as in every other property, and that consequently, even in these cases of great similarity, Haüy's law is strictly true. The only cases excluded are cubic crystals, for which the symmetry fixes the angles. But even in these cases all the other properties show progressive differences.

§ (14) Analysis of Crystal Structure by X-rays (Laue).—At this point the writer was eagerly looking for a means of converting into absolute measures these relative dimensions of the space-lattice cells, afforded by the space-ratios. For so far only strictly related compounds, such as those of isomorphous series for which the structures were certainly analogous, were strictly comparable. One step forward had been made, however, for from the evidence of the production of excellent mixed crystals and parallel growths it had been proved that rubidium and ammonium sulphates were also strictly comparable, and their isostructure rendered certain; for closeness of structural dimensions is a condition for such productions. But at this moment, in the year 1912, M. von Laue of Munich made his epoch-making suggestion, that if the atoms or molecules of a crystal are in truth arranged in a space-lattice, they ought to be capable of reflecting or diffracting the exceedingly minute electro-magnetic waves of the X-rays, for the order of dimensions of atoms and of the wave-lengths of X-rays is approximately the same, about 10^{-8} cm. The planes of atoms of such a space-lattice ought, in short, to behave much as a diffraction grating does towards ordinary rays of light. His colleagues, W. Friedrich and F. Knipping, tried the experiment with a crystal of zinc blende, and found it to succeed perfectly, the crystal diffracting the incident beam of X-radiation into a number of diverted beams, each of which made an impression on a photographic plate which developed as a spot, and the whole series of spots formed a pattern—the now well-known Laue radiogram—which exhibited the cubic symmetry of the crystal, each plane of atoms of the space-lattice producing its own spot. The writer was fortunate in seeing these historic first Laue radiograms in the Munich laboratory on a visit just afterwards. One of them, for zinc blende, is reproduced in *Fig.* 27.

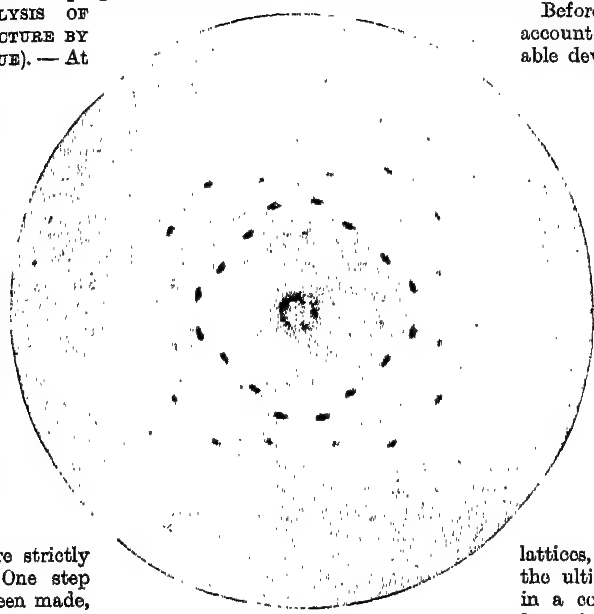


Fig. 27.—Reproduction of original Laue Radiogram of Zinc Blende.

Before giving a brief account of the remarkable development which has followed from this new mode of attack, it may be said at once that the results have fully confirmed the conclusions of crystallographers above described. They prove beyond a shadow of doubt that crystals are built up on the principle of point-systems and space-lattices, the atoms being the ultimate units, and in a considerable number of cases of simple chemical compounds the absolute dimensions of the space-lattice cells have been determined. In the particularly important case of the sulphates of potassium, rubidium, and caesium, for which the relative cell dimensions had been given by the writer as explained in a previous section, the absolute values now obtained by X-ray analysis have proved to possess precisely these relations, the correspondence being of a most satisfactory and even surprising closeness. The actual figures are given and more fully referred to in § (17).

§ (15) The X-ray Spectrometer (Bragg).—Shortly after the publication of the memoir of Laue, Friedrich, and Knipping, a new method of procedure[1] was devised by Sir William H. Bragg, and a fuller explanation was given of the Laue radiograms by his son, Prof. W. L. Bragg, who has also used with

[1] See "X-Rays," §§ (6)-(9), Vol. II.

conspicuous success the method of his father. This consists in mounting the crystal on a spectrometer-goniometer, replacing the ordinary collimator by two successive slits in leaden screens (lead being impervious to X-rays), and the telescope by an ionisation tube, the diffracted or reflected X-rays passing within the latter and ionising the vapour of methyl bromide or gaseous sulphur dioxide contained therein, the electrical effect being communicated to an electroscope and measured as to its intensity. It was found that reflection of the X-rays at any plane of atoms (really a large number of parallel planes inside the crystal having the same relations to the crystal symmetry) occurred only at certain specific glancing angles, the facts being expressed by the following equation, which is of similar kind to that governing the action of a diffraction grating toward light waves:

$$n\lambda = 2d \sin \theta,$$

where n is the order of X-ray spectrum, λ the wave-length of the X-rays (the Braggs using "monochromatic" X-rays instead of the more general X-radiation employed by Laue and his colleagues), d is the "spacing" of (distance between) the parallel planes of atoms affording the reflection, and θ is the glancing angle of reflection (that is, the angle from the plane, not from the normal to the plane). Knowing the wave-length of the X-rays employed—and Sir William Bragg had determined this with considerable accuracy for the rays from certain specific anticathodes, notably those of palladium—it is obvious that this important equation affords us, in the value of d, the absolute distance between the planes of atoms, and between the atoms themselves when planes normal to all three directions of space are investigated; indeed it gives us the absolute lengths of the edges of the space-lattice cells.

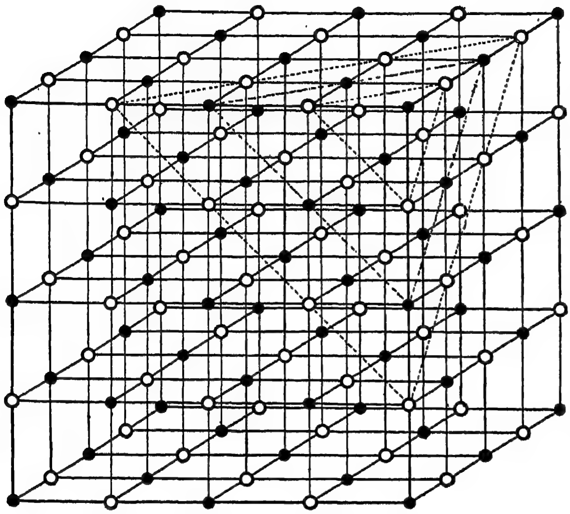


FIG. 28.—The Structure of Sylvine and Rock-salt.

Really satisfactory results, consisting of a definite determination of the positions of all the atoms, have only so far been obtained with substances of relatively simple chemical composition. They are divisible into two very distinct types: (1) those in which the positions of all the atoms are fixed by the symmetry, by the nature of the space-lattices or other point-systems present, which are clearly recognised by the X-ray analysis; and (2) those in which only the atoms of one or more of the dominating elements present—the metal, for instance, in a salt or binary compound—are thus fixed, while the atoms of another element or other elements are permitted some latitude within certain limits, the exact positions being determinable by the X-ray analysis. Three instances of the former fixed type are: (*a*) many native metals, copper for instance, the structure of which is simply that of the face-centred cube lattice already illustrated in *Fig.* 26; (*b*) potassium or sodium chloride, KCl (sylvine) or NaCl (rock-salt), shown in *Fig.* 28; and (*c*) zinc blende, ZnS, portrayed in *Fig.* 29.

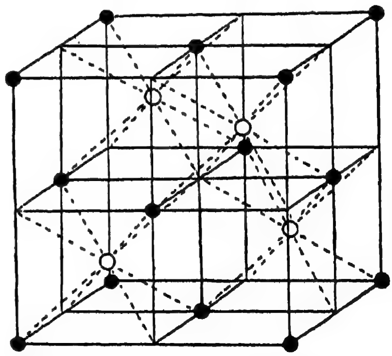


FIG. 29.—The Structure of Zinc Blende.

Two instances of the latter type, with one or more variables, are iron pyrites, FeS_2, and calcium carbonate, $CaCO_3$, the positions of the sulphur and oxygen atoms being allowed a certain choice of position along particular lines; these positions have been exactly determined by quantitative X-ray analysis.

The structure of iron pyrites is shown in *Fig.* 30, which is the reproduction of a photograph of a model. The case of calcium carbonate is more complicated, as the structure appears to be built up of calcium atoms and of CO_3 groups, in which the oxygen and carbon atoms

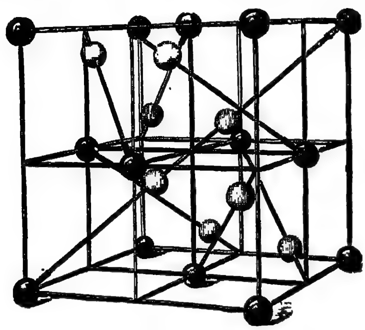


FIG. 30.—The Structure of Iron Pyrites.

are very closely bound together, for a reason to be presently explained (the sharing of certain electrons in common).

In the case of copper crystals the atoms are all alike, being those of the element copper, and their arrangement is that of the face-centred-cube space-lattice. In the cases of the chlorides of sodium and potassium, however, two kinds of atoms, those of the alkali metal and of chlorine, are present. Now the Bragg method has revealed another important and very helpful fact, that the intensity of X-ray reflection from a crystal face (really from the planes of atoms within the crystal parallel to the face) is proportional to the mass of the atoms composing it, and the atomic weight or atomic number may be taken as representing the mass of the atom. Hence, as potassium possesses the atomic weight 39 and chlorine 35·5 the masses of the atoms in sylvine are nearly alike. But as sodium has the much lower atomic weight, 23, there is a distinct dissimilarity in the masses of the two kinds of atoms composing rock-salt. The effect of this is curious. Referring to *Fig.* 28, it will be seen that the cube planes (100) are all composed of equal numbers of metallic and chlorine atoms, while the octahedral planes (111), shown in dotted lines, are alternately composed all of chlorine and all of metallic atoms. The effect in the case of rock-salt is as if there were two interpenetrating space-lattices of the face-centred cube, of different reflecting power, composed solely of sodium and of chlorine respectively, each of double the spacing; the first order spectrum becomes consequently weakened, while the second order is abnormally strong, the third order nearly disappears, but the fourth order reflection is quite good. On the other hand potassium chloride (sylvine) crystals behave as if the structure were of one kind of atom only, and the space-lattice is that of the simple cube, the small cubes of *Fig.* 28.

§ (16) DEBYE AND SCHERRER.—Yet a third method of X-ray analysis has been devised and most successfully used by P. Debye and P. Scherrer, and independently by A. W. Hull, in which the crystal is pulverised and the powder compressed into a rod, placed in the axis of a cylindrical photographic film and subjected to "monochromatic" X-rays. Characteristic interference curves are shown on development of the film, afforded by such particles (and among the infinite variety of orientations of the particles some such are bound to be present) as are correctly orientated to give the reflections from their planes of atoms. These three methods, the Laue radiographic, the Bragg spectrometric, and the Debye and Scherrer powder-method, of X-ray analysis, most wonderfully confirm and supplement one another.

Sir William Bragg has since shown that the powder method may be adapted for use with the X-ray spectrometer, it being only necessary to paste the powder on the flat surface of a holder placed on the spectrometer instead of the crystal.

§ (17) STRUCTURE AS DETERMINED BY X-RAYS.—Careful examination of the reflections of X-rays at the various glancing angles corresponding to the different orders of spectra (n), as regards both the determination of the exact angle and of the relative intensity, has enabled the structure of a large number of substances of more or less simple character to be determined and their cell dimensions to be measured.

(i.) *Zinc Blende, Diamond, and Fluorspar.* —That of zinc blende, ZnS, is shown in *Fig.* 29. The zinc atoms occupy the corners and centres of the large cube faces, but the sulphur atoms occupy the centres of alternate small cubes. Moreover, if we replace both zinc and sulphur by carbon atoms we have the structure of the diamond. Again, if the zinc atoms be replaced by calcium atoms, and the centres of all the cubelets (not only alternate ones) be filled with fluorine atoms, we have the structure of fluorspar, CaF_2. In all these cases the structure-lattice itself fixes the positions of all the atoms.

(ii.) *Iron Pyrites.* — In the case of iron pyrites, shown in *Fig.* 30, only the iron atoms are fixed by the lattice itself, of which they form the corners, like the zinc and calcium atoms of zinc blende and fluorspar. The sulphur atoms, however, in iron pyrites are not at the centres of the cubelets, but each is moved along one of the diagonals (a trigonal axis), chosen complementarily as shown in

the figure, to a position the exact location of which has been determined with great accuracy; as no iron atom is present at the corner near which the sulphur atom is arranged, its next neighbour will be another sulphur atom, as shown about the central corner in *Fig.* 30.

(iii.) *Carbon and its Compounds.*—The structure of diamond has already been given in (i.); that of the other form of carbon, graphite, has been determined by the powder method, by Debye and Scherrer, by Hull, and by Sir William Bragg, and found to be, as expected, that of a trigonal space-lattice. Now it is particularly interesting that the two structures, of diamond and of graphite, correspond to two different arrangements of the four valency bonds of carbon. Those of diamond (cubic) are arranged tetrahedrally, each carbon atom being attached to four others situated at the corners of a tetrahedron (of which it forms the centre) resting on one of its faces, one bond being thus upright and the other three radiating and slanting downwards like a tripod. In graphite there are three principal valencies in a plane, which is puckered according to Hull, and a fourth feebler one perpendicular to the plane. Further, Sir William Bragg has shown that the same puckered planes exist in both diamond and graphite, and that their repetition in the structure results in the formation of six-carbon-atom rings. Moreover, if we take two such layers of a diamond structure model and remove one of them further away from the other while at the same time giving it some rotation, we produce a model of graphite. The greater separation of the planes is also accompanied by some tightening up of the atoms in the puckered plane, and the two occurrences together determine that graphite has great cohesion in the plane and very slight at right angles to the plane; hence, it cleaves parallel to the plane so readily as to be actually soft enough to act as a lubricant. On the other hand, diamond is the hardest substance known.

Most interesting of all, however, is the fact brought to light by Sir William Bragg, that these hexagonal six-atom rings persist as such in the aromatic carbon compounds. Benzene itself, C_6H_6, has not yet been investigated as the crystals melt at 6° C. But there is good ground for believing that it consists entirely of such hexagonal carbon rings with attached hydrogen atoms. Naphthalene, however, $C_{10}H_8$, which has two benzene rings in its constitution, two carbon atoms being common to the two rings, crystallises well, and anthracene, $C_{14}H_{10}$, well enough for the powder method, this hydrocarbon having three benzene rings with four carbon atoms in common. On analysis by the X-ray spectrometric method it was found that the monoclinic space-lattice cells of each substance contained two molecules of the hydrocarbon, and their absolute dimensions were determined. The corners of the cells, and also the centres of their basal plane faces, are each occupied by the double benzene ring of naphthalene or the triple one of anthracene, each acting as an entity, the hydrogen atoms being accommodated in the spaces left around these structures. Further, the double benzene rings were also found to remain intact in the derivatives of naphthalene, the particular ones studied being acenaphthene, the α- and β-naphthols, and α-naphthylamine. The cell dimensions of these substances were determined in absolute measure, and the cells found to contain four molecules of the substance in all four cases.

In these remarkable organic structures it is clear that the molecules persist in the crystalline state, and a clear case is afforded which should check the tendency to conclude, from the early X-ray results with very simple binary compounds, that atoms alone need be considered in the crystalline state, and that molecules no longer persist. Such a conclusion was in any case premature, and there are many reasons why it was difficult to believe. For saturated—indeed supersaturated—solutions, and not dilute, are concerned in crystallisation, so that ionic dissociation is excluded; and the molecules which deposit themselves in orderly fashion in forming the crystal are reproduced, or others indistinguishable from them, when the crystal edifice is taken down again by solution or fusion.

(iv.) *Alkali Sulphates.*—As an example coming under the writer's own personal observation, of the highly satisfactory manner in which these results confirm the work of crystallographers, the case of the rhombic alkali sulphates may be quoted. The sulphur atoms were found to be located at the corners and face-centres of the unit rectangular prism, practically like *Fig.* 28, except that the rectangular spacings (edges of the prism-cell) were not equal, the block or cell being not cubic but rectangular rhombic. The metallic atoms, of potassium, rubidium, or caesium, are arranged hexagonally, which is in remarkable accordance with the well-known closeness of potassium sulphate and its isomorphs to hexagonal symmetry, the angles in the prism zone being only a few minutes removed from exactly 60° and 30°. The table on following page shows the writer's published values for the molecular volumes and topic axial ratios (the relative dimensions of the space-lattice cells) for the four salts of the series, and also the absolute dimensions and volumes as determined by means of X-rays with the same crystals (grown and measured by the writer) by Prof. Ogg and

Mr. F. L. Hopwood in Sir William Bragg's laboratory. The correspondence is striking, indeed if the writer's relative values be multiplied by 10^{-7} and considered as centimetres the absolute values as fixed by X-ray reflection are practically reproduced. Moreover, the almost perfect isostructure (equality of cell dimensions) of the rubidium and ammonium salts is fully confirmed. Incidentally, this is a fact which clearly shows the fallacy of the Pope and Barlow valency volume theory, which would require that the volume of the ammonium salt should be twice as great as that of rubidium sulphate, the respective valency volumes of the two salts being 24 and 12.

a periodic one like that just referred to, with the electro-positive alkali metals at the summits of the periods and the halogens and other electro-negative elements at the depressions; it is shown in *Fig.* 31. The sphere of which this "atomic diameter" is the diameter is the limit of approach to that of any other atom, and in the simplest structures, the crystals of the elements themselves, it is the actual distance of separation of the centres of the atoms from one another. The law is proving very helpful in elucidating further more complicated crystal structures; for no two atoms can occupy, or be assigned, closer positions than the sum of their two radii.

COMPARISON OF MOLECULAR VOLUMES AND SPACE-RATIOS WITH ABSOLUTE DIMENSIONS OF SPACE-LATTICE

Salt.	Molecular Volume.	Topic Axial Ratios. ψ for $K_2SO_4=1$. $\chi : \psi : \omega$	Absolute Lengths of Sides of Unit Rhomb. a	b	c	Volume of Unit Rhomb.
			cm.	cm.	cm.	c.c.
K_2SO_4 .	64·91	0·5727 : 1·0000 : 0·7418	$5{\cdot}731 \times 10^{-8}$	$10{\cdot}008 \times 10^{-8}$	$7{\cdot}424 \times 10^{-8}$	$425{\cdot}78 \times 10^{-24}$
Rb_2SO_4 .	73·34	0·5944 : 1·0387 : 0·7774	$5{\cdot}949 \times 10^{-8}$	$10{\cdot}394 \times 10^{-8}$	$7{\cdot}780 \times 10^{-8}$	$481{\cdot}14 \times 10^{-24}$
$(NH_4)_2SO_4$	74·04	0·5946 : 1·0552 : 0·7723	$5{\cdot}951 \times 10^{-8}$	$10{\cdot}560 \times 10^{-8}$	$7{\cdot}729 \times 10^{-8}$	$485{\cdot}71 \times 10^{-24}$
Cs_2SO_4 .	84·58	0·6213 : 1·0877 : 0·8191	$6{\cdot}218 \times 10^{-8}$	$10{\cdot}884 \times 10^{-8}$	$8{\cdot}198 \times 10^{-8}$	$554{\cdot}88 \times 10^{-24}$

§ (18) THE LAW OF ATOMIC DIAMETERS.—Another remarkable principle has more recently been deduced by Prof. W. L. Bragg from a consideration of all the X-ray analyses obtained up to the beginning of the year 1920. It is that the atoms of each chemical element possess the same diameter (regarding the atom as a sphere) in all the compounds of that element studied, and these fixed and definite diameters are related, not as the valencies of the elements (as they are supposed to be in the Pope and Barlow theory), but in a manner reminiscent of the relations of their atomic volumes, as shown in the well-known curve of Lothar Meyer. The curve is

§ (19) X-RAY CRYSTAL ANALYSIS AND ATOMIC STRUCTURE.—Finally, X-ray analysis in the hands of the Braggs is now affording some indication also of the structure of the atoms composing the crystal structure. It appears quite likely that the sphere just referred to as representing the limiting surface of the atomic domain is the outer shell of electrons. More careful study of the intensities of the reflections of the various orders of X-ray spectra from the diamond (from the tetrahedron planes of atoms), and from fluor-spar, are affording indications of something—probably connecting electrons—dispersed tetrahedrally in the case of the diamond, and

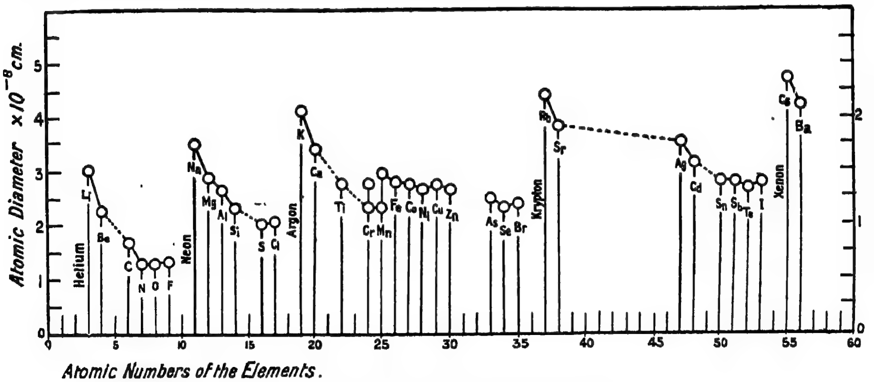


FIG. 31.—The Curve of Atomic Diameters.

one-third to one-quarter of the way between the calcium and fluorine planes in the case of fluorspar. Also by using X-rays rendered more strictly monochromatic by preliminary reflection from rock-salt, much more accurate determinations of intensity have been obtained, such indeed as can be expressed by a definite mathematical formula. Among the various factors which make up this formula there is one which relates to the positions of the electrons of the outer shell, and this again assists in obtaining indications of the positions of these electrons. Hence, it is probable that further researches will afford us not only accurate X-ray analysis of crystal structure, but also of atomic structure.

The results so far are more in unison with the Lewis-Langmuir version of the theory of atomic structure than with the Bohr-Sommerfeld variety. There is distinct evidence of the sharing of electrons by the electro-negative elements, and in cases of the combination of an electro-positive element with an electro-negative one, such as that of potassium and chlorine to form crystals of sylvine, KCl, that the one (here the potassium atom) loses an electron, while the other (here the chlorine atom) takes it up. In each of these two cases the more stable outer shell of the argon type with 18 electrons appears to be formed with each element, the number of electrons of these two elements, potassium and chlorine, being 19 and 17 respectively, corresponding to their atomic numbers in accordance with the immensely important law of Moseley. The unit excess of positive charge left on the nucleus of the potassium atom, and the negative excess unit charge of the added electron on the chlorine atom, act as the attractive connecting force binding the two atoms together.

§ (20) Crystallographic Significance of Moseley's Law.—The law of Moseley just referred to has also a profound crystallographic significance. This law, which it will be remembered emerged from Moseley's last two memoirs on "High Frequency Spectra" (he was killed at Suvla Bay, Dardanelles, 1915), is the essence of atomic structure, as now generally agreed, and is not dependent on the special features of either the Bohr or the Langmuir versions. It stipulates that the positive charge on the nucleus is N units, where N is the atomic number, the sequence number of the element in the periodic table, and that there are N electrons, each of unit negative charge, surrounding it, to counterbalance the nucleus and form the atom. According to Langmuir these electrons are arranged in successive shells containing 2, 8, 8, 18, 18, and 32 electrons as we proceed along the periodic table, while according to Bohr the shells comprise successively 2, 8, 18, 32, 18, and 8 electrons. Which of the two versions is correct is for the moment immaterial as regards the point now desired to be emphasised. This is that the progression in complexity of the atoms provided by Moseley's law explains most fully and satisfactorily the progression in the crystallographic elements, angles, space-lattice cell-dimensions, and in the optical and other physical properties, which has been observed by the writer to occur in the two large and important rhombic and monoclinic series of isomorphous salts (sulphates and selenates, and double sulphates and selenates with $6H_2O$) containing the alkali metals, potassium, rubidium, and caesium, when potassium is replaced by rubidium and the latter in turn by caesium. For it has been shown that these strongly electro-positive elements form the summits of the Bragg curve of atomic diameters, as clearly shown in *Fig.* 31, and their atoms increase in complexity by a complete shell of electrons at each step (from potassium to rubidium, and from rubidium to caesium), and their atomic numbers are 19, 37, and 55, differing by equal steps of 18. Hence, this regular addition in mass and complexity of the atom at each step is bound to produce a corresponding crystallographic progression, such as has been so fully revealed by the writer's work of many years, and which has now been so strikingly confirmed by the direct absolute measurement of the space-lattice cells by means of X-rays.

§ (21) Optical Properties of Crystals.—The physical properties of crystals—their Optics, their Thermal, Elastic, Electrical, and Magnetic Properties—are of equal importance and interest to the Morphological characters which have been considered in the preceding pages. For all of them are profoundly affected by the symmetry, inasmuch as this symmetry is not only that of the exterior form but of the internal structure. Indeed, it has already been made clear that the former is but the natural expression of the latter.

The optical properties[1] are of prime importance as being easily observed and often characteristic of the particular substance, and also because they frequently afford the means of readily deciding as to the type of symmetry, when the goniometrical examination leaves a doubt between certain alternatives or when crystals of adequately perfect exterior form are unprocurable. Three main operations are in general involved in the optical investigation, and they are all connected with the central fact that the optical properties of a crystal may be represented by an ellipsoid of general form, that is, one of which the three rectangular principal axes are unequal in length. Two varieties of the ellipsoid have been used, the vibration-velocity ellipsoid of Fresnel, and its

[1] See "Light, Double Refraction of"; also "Polarised Light," §§ (5)-(16), etc.

polar reciprocal, the indicatrix of Fletcher. The latter is the simplest and most useful for all practical purposes, as its axes are directly represented by (proportional to) the principal optical constant, the refractive index, corresponding to light vibrating along each of those three important directions. Hence, the main task in practical crystal-optics is to determine the refractive index of the crystal in the three rectangular directions corresponding to the axes of the optical indicatrix. The determination of refractive index is an operation familiar in the physical laboratory, and it is only necessary here to mention that it involves the measurement of the angle of the prism (on the goniometer, like a crystal angle), and that of minimum deviation for light of a number of wave-lengths distributed through the spectrum, the same goniometer-spectrometer being used for both operations.

We have first, however, to know these three axial directions of the indicatrix, and if they are not already fixed by the development of high symmetry on the crystal, to determine them, the determination of two sufficing to fix all three. This is achieved by examination of the phenomena afforded by plates of the crystal in parallel and convergent polarised light.

Now the greater the symmetry present the simpler this task becomes. If the crystal possess the perfect cubic symmetry the optical properties are the same in all directions within the crystal, and the ellipsoid becomes a sphere; a 60° prism cut out of the crystal in any direction will, therefore, give us the refractive index, just as if it were glass. There is consequently for a cubic crystal but one refractive index for light of any one wave-length. If the symmetry be that of the tetragonal, hexagonal, or trigonal system the ellipsoid becomes one of rotation about the tetragonal, hexagonal, or trigonal axis of symmetry. The refractive index for light vibrating along that axis will thus be different from that vibrating in any other direction, the maximum difference being for all rays vibrating in the circular section of the ellipsoid perpendicular to the axis. These two extreme refractive indices are respectively labelled ϵ (vibrations parallel axis) and ω (vibrations perpendicular axis and in circular section). If ω be the greater, as for calcite $CaCO_3$, the crystal is conventionally negative, but if ϵ have the higher value, as for quartz, SiO_2, the crystal is said to be positive. Light travelling along the axis remains a single beam, as its vibrations are all in the circular section. But light travelling in any other direction is doubly refracted, a bifurcation into two rays occurring, one of which always affords the index ω and is an ordinary ray, while the other is an extraordinary ray which only affords ϵ when its vibrations occur parallel to the axis. Hence, tetragonal, hexagonal, or trigonal crystals are optically "uniaxial."

In actual practice a 60° prism is cut so that the refracting edge is parallel to the tetragonal, hexagonal, or trigonal axis; this prism affords immediately ϵ and ω; for the ray traversing it at minimum deviation passes along a diameter of the circular section, and divides into two rays vibrating perpendicularly to its path and polarised at right angles to each other, one of which vibrates parallel to the refracting edge and axis and affords ϵ, while the other vibrates in the circular section and affords ω. The two images of the signal-slit corresponding to these two rays can readily be distinguished by means of a Nicol prism placed in front of the telescope eye-piece of the goniometer-spectrometer.

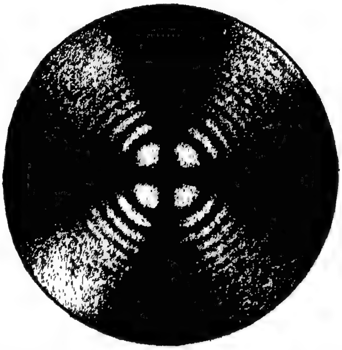


Fig. 32.—Uniaxial Interference Figure in Convergent Polarised Light.

There is usually no difficulty in recognising a uniaxial crystal, and finding the position of its axis. For a plate cut perpendicularly to the latter exhibits in convergent polarised light the well-known interference figure consisting of circular iris-coloured rings and a co-central black cross, the former being curves of equal retardation of one of the two rays behind the other, and the cross marking the directions of vibration of the crossed Nicols. The goniometry will have indicated the direction which is probably that of the tetragonal, hexagonal, or trigonal axis, and a plate cut or ground from the crystal perpendicularly to this direction can readily be tested for the production of this unmistakable uniaxial interference figure, a photographic reproduction of which is given in *Fig.* 32.

The more general case of an optical ellipsoid with three unequal rectangular principal axes corresponds to rhombic, monoclinic, or triclinic symmetry. The directions of the three axes are identical with those of the rhombic axes, but only one crystallographic axis, the symmetry axis *b*, is coincident with an axis of the ellipsoid of a monoclinic crystal, and in the case of a triclinic crystal there are no coincidences of morphological and optical axial directions. In order to determine the

three refractive indices α, β, γ, corresponding to vibrations parallel to the three respective axes of the ellipsoid, we can proceed at once in the case of a rhombic crystal to cut or grind and polish three 60° prisms, so that the refracting edge of each is parallel to one of the crystallographic axial directions (a different one in each case), and so that the bisecting plane of the prism is parallel to a principal plane of the ellipsoid; this plane will then also contain a second axis of the ellipsoid. Hence, such a prism affords two images of the signal-slit of the spectrometer, corresponding to two of the refractive indices, say α and β, or β and γ, or α and γ. Two of the prisms suffice to afford us all three indices, and one in duplicate; but the three prisms give each index twice over, affording an excellent test of accuracy. Each index is, of course, determined for light of a convenient series of wave-lengths distributed over the spectrum.

a general triaxial ellipsoid must have two radii lying in its principal section-plane (that containing the α and γ axes) which are equal to the intermediate axis β perpendicular to that plane, and so has two circular sections perpendicular to which there will apparently be equal (no double) refraction and the property of an optic axis exhibited. Hence rhombic, monoclinic, and triclinic crystals are "biaxial," and if a plate be cut perpendicular to that axis of the ellipsoid which is the bisectrix of this acute optic axial angle, we shall see (in the dark field in convergent polarised light) the well-known biaxial figure of iris-coloured lemniscates (rings, loops, and ellipse-like curves) and dark extinction brushes, as shown in *Fig.* 33. The two optic axes are at the central points of the two systems of rings, and are further indicated by the fine vertices of the hyperbolic brushes, which pivot

Fig. 33.—Biaxial Interference Figure in Convergent Polarised Light.

Fig. 34.—Biaxial Interference Figure with Nicols rotated simultaneously 45°.

In the case of a monoclinic crystal we have first to determine the situation of the two rectangular axes of the ellipsoid which lie in the symmetry plane. We do this by cutting or grinding a plate parallel to the symmetry plane, leaving the faces along its edge undamaged so as to serve as reference faces. This is then placed in the polariscope arranged for parallel light, and the directions determined (with respect to one of the reference faces) for which the dark field of the crossed Nicols is reproduced. These two perpendicular directions are those of the required axes of the ellipsoid. We then know all three axes of the indicatrix and can proceed to prepare the three prisms as in the case of a rhombic crystal.

The case of a triclinic crystal is more difficult, but the sequence of operations is the same; the section-plates required to fix the directions of the axes of the ellipsoid by extinction determinations need, however, to be more numerous.

We then proceed to study the interference phenomena in convergent polarised light. For

about the optic axes when the plate is rotated in its own plane, or when the two Nicols are simultaneously rotated instead, as shown in *Fig.* 34 for 45° of rotation. The angle between the optic axes as thus seen in air is not the true angle within the crystal, but this latter can be determined by preparing a second section-plate perpendicular to the obtuse bisectrix, and measuring the apparent optic axial angle of both plates in one and the same highly refractive liquid, when the following simple formula enables us to calculate the true angle $2V_a$, within the crystal, from the apparent acute and obtuse angles $2H_a$ and $2H_o$ in the liquid: $\tan V_a = \sin H_a / \sin H_o$.

These measurements must also be made for monochromatic light of the same series of wave-lengths as the refractive indices; for in general the angle is dispersed like a spectrum, and is different for each wave-length, giving rise to the chromatic effects seen in white light. Indeed, this dispersion is sometimes so great that for red light the optic axes are separated in another (perpendicular) plane to that

which contains them for blue light, and very beautiful interference figures are afforded in such cases of "crossed-axial-plane dispersion." The writer's spectroscopic monochromatic illuminator, described to the Royal Society in the year 1894, which supplies monochromatic light of a purity equal to the three-hundredth part of the spectrum, enables the phenomena to be very accurately followed, and the exact wave-length for the crossing, when the uniaxial figure is temporarily produced, to be determined.

and the truly orientated plates and prisms can be prepared with the aid of the cutting and grinding goniometer which the writer also described to the Royal Society at the same time as the illuminator, and which has proved so invaluable that no excuse is needed for illustrating it in *Fig.* 35. It possesses all the attributes of a goniometer, having in addition its adjusting arcs divided, combined with a cutting apparatus (revolving soft iron disc with diamond-fed edge) and a grinding and polishing lap (one of ten interchangeable ones of various suitable materials). Either cutter or grinder are removable, to leave the field clear for the other, and a pressure-controlling apparatus is also provided which effectually prevents fracture of the crystal. There are also labour-saving devices for preparing a second surface parallel to a first, or at 60° (or any other definite angle) to the first. With this instrument the many hundreds of section-plates and 60° prisms used in the writer's researches have been prepared, and the accuracy of the results is largely due to it.

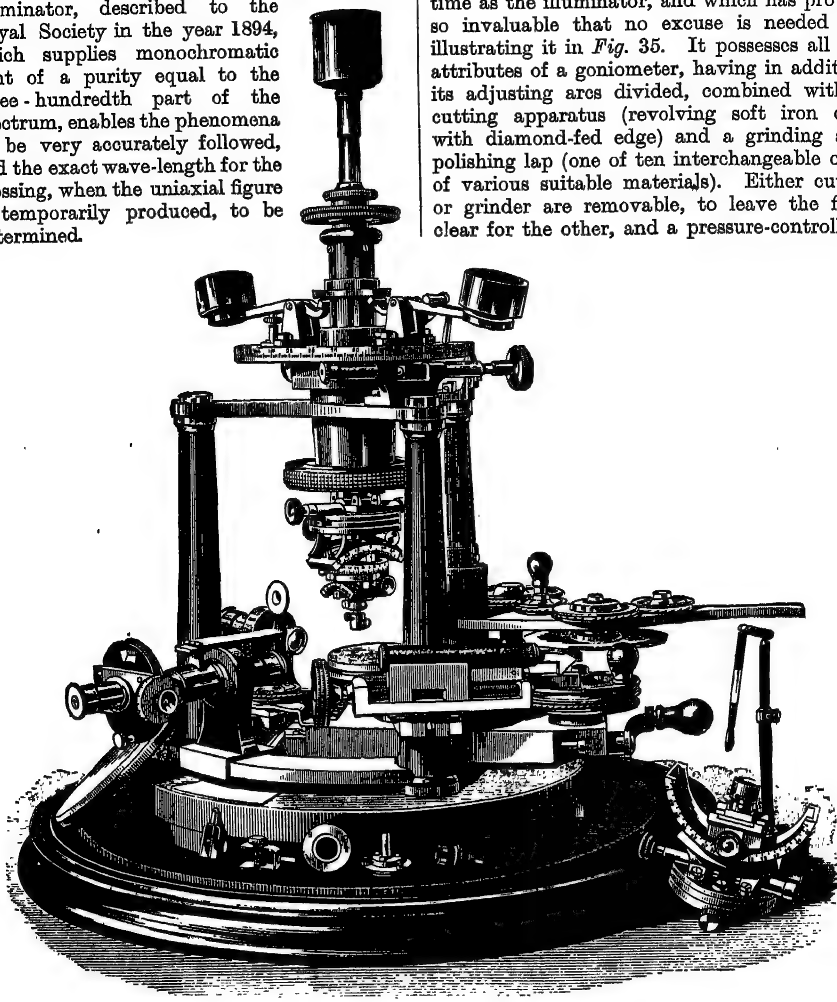


FIG. 35.—The Cutting and Grinding Goniometer.

It will now be clear that the optic axial phenomena afforded by a crystallised substance are always specific and very often characteristic of the crystals of the substance. It will be obvious, however, that a ready mode of producing pure monochromatic light of any wave-length is essential, and also a means of preparing parallel faced and truly plane section-plates and 60° prisms of any required orientation as regards the crystal faces and structure. The former is very conveniently afforded by the apparatus just referred to,

§ (22) Other Physical Properties of Crystals.—Just as the symmetry of a crystal imposes its determinative and controlling effect

on the optical properties of the crystal, so does it also in like manner determine the orientative character of the thermal, electric, magnetic, and elastic properties. The limits of this article have been reached and these properties will be found more or less dealt with under other articles of this Dictionary. But from the practical point of view it should be stated here that as crystals are usually such small objects the most refined methods of measurement are required, the ordinary large-scale methods of the physical laboratory being quite inapplicable. The wave-length interferometer is therefore the indispensable instrument of measurement. Indeed, it was in order to measure the thermal expansion of crystals that Fizeau devised this interferometric method, using curved interference fringes in sodium light as the unit of the scale. The difference of expansion of the screw-legs of a platinum tripod carrying a plano-convex lens, and of the crystal resting below the lens on the table (through which the three screws forming the tripod pass) of the tripod, was the actual object of the measurement. The method has been brought to greater perfection in the writer's interference dilatometer, in which half-wave-length rectilinear interference bands in the more purely monochromatic red hydrogen, red cadmium, or yellow neon light (the two latter being absolutely monochromatic, affording none of the secondary interference in periods, so disconcerting with sodium light) are used as the coarse units; and as these are suitable for use with a micrometer, the hundredth part of such a unit is readily measurable. Thus the one-eight-millionth of an inch, or the one-three-thousandth of a millimetre, is the fine unit. A platinum-iridium tripod carries a glass plate (not a lens) to furnish one of the surfaces (that resting on the tops of the three tripod-forming screws) reflecting the interfering light, the crystal upper surface itself, or the surface of an aluminium or black glass disc carried by it furnishing the other. (Aluminium is especially suitable, as by its large expansion a stout disc nearly compensates for the expansion of the but slightly expanding platinum-iridium screws.)

An adaptation of the same principle, and part of the same optical apparatus, to the determination of the elastic bending of a plate or bar of the crystal is also used in the writer's elasmometer, which confers the same accuracy on the determination of the modulus of elasticity of crystals. As in order to complete the elastic constants of a crystal torsion determinations are also required, a torsion-meter has also now been constructed and installed in the author's laboratory, for the application of this same refined interferometric method to the much more difficult task of determining the modulus of torsion of crystals. Moreover, this new instrument is an interferometer of general application; for it possesses all the essential features of the interference comparator for standards of length, constructed under the writer's supervision for the Standards Department of the Board of Trade, and will enable the determination to be carried out of any very minute movement or short distance in general, such, for instance, as the very small piezo-electrical movements of crystals which are so much in need of further study. A very large field of wonderfully steady black interference bands on the coloured ground of the monochromatic light employed is afforded, and one of the reflecting surfaces producing the interference is actually carried by the observing travelling microscope, the lateral movement of which over a V-and plane bed is effected by an extremely fine screw, and so steadily that the interference bands traverse the field parallel to the pair of vertical spider-lines without the slightest flicker, and can be held at any position for any length of time at the will of the observer.

Full details of all these instruments for crystallographic research (and for much more general physical purposes at the same time) will be found in the writer's *Crystallography and Practical Crystal Measurement* (Macmillan & Co.), vol. i. of the 2nd (1922) edition of which deals with crystal morphology, and vol. ii. with the physical properties of crystals.

In conclusion it may be emphasised that in all these refined measurements it has been fully confirmed that the internal structural symmetry of crystals, which is so beautifully exhibited in their exterior form, rules also absolutely over even the minutest details of their physical properties. For this reason the study of crystals should appeal more and more in the future to physicists, and a knowledge of Crystallography has now become one of the most valuable aids to original investigation in Physics.

A. E. H. T.

CRYSTALS, Optical Properties of. See "Crystallography," § (21).

Physical Properties of: thermal, electric, magnetic, and elastic. See *ibid.* § (22).

Relation between morphological and optical properties. See "Polarised Light and its Applications," §§ (7) (iv.) and (16).

The Seven Systems of. See "Crystallography," § (7).

CURVATURE OF FIELD OF AN OPTICAL INSTRUMENT. One of the five third-order aberrations of a lens. See "Lens, Theory of Simple," § (11); also "Telescope," § (3).

CUT GLASS, PREPARATION OF. See "Glass," § (36).

CZUDNOCHOWSKI, VON, PHOTOMETER. See "Photometry and Illumination," § (108).

— D —

Damping Error of Gyro Compass. See "Navigation and Navigational Instruments," § (15).

Daylight Attachment for Illumination Photometers. See "Photometry and Illumination," § (76).

Daylight Factor. See "Photometry and Illumination," § (75).

Daylight Illumination. See "Photometry and Illumination," § (74) *et seqq.*

Dead Reckoning: method of determining a ship's position. See "Navigation and Navigational Instruments," § (8).

Debye and Scherrer's method of X-ray analysis, in which the crystal is pulverised and the powder compressed into a rod, placed in the axis of a cylindrical photographic film and subjected to "monochromatic" X-rays. Characteristic interference curves are shown on development of the film. See "Crystallography," § (17).

Decomposition of Glass, caused by chemical re-agents. See "Glass, Chemical Decomposition of," § (2).

Density of Glass. See "Glass," § (22).

Detail Surveying, general methods. See "Surveying and Surveying Instruments," § (15).

Deviation of Compass, method of correction. See "Navigation and Navigational Instruments," § (12).

Devitrification: destruction of the vitreous character of glass, by partial or complete crystallisation, generally of silicates. See "Glass," § (20); also "Glass, Chemical Decomposition of," § (1).

Diaphragm Test, Bishop Harman's, a popular form of photometer used in the British Army. See "Ophthalmic Optical Apparatus," § (3).

Diffraction: the term used to denote the departures from the law of linear propagation which occur when light waves pass an obstacle of any character. For the effects of diffraction in the production of grating spectra see "Diffraction Gratings, Theory of," §§ (2) and (4).

Diffraction: effect on theory of telescopes. See "Telescope," § (7).

Diffraction: effect of, in vision. See "Eye," § (23).

Diffraction Grating, theory and use of, in wave-length measurements. See "Wave-lengths, The Measurement of," § (2).

DIFFRACTION GRATINGS, THE MANUFACTURE AND TESTING OF

Introduction.—The ideal grating consists of a large number of grooves or "lines" on an optical surface. The grooves should all be exactly alike; and if the surface is plane, they must also be exactly straight, parallel, and equidistant. If the surface is spherical the grooves, as projected on the tangent plane to the sphere at the centre of the grating, should be straight, parallel, and equidistant.

It is needless to say that the ideal grating has never been realised, and never will be. The best existing gratings do, however, approach it very nearly—the degree of approximation being about the same as that of a good telescope or microscope objective to the ideal objective.

It is the purpose of the present article to give an account of the manufacture of gratings, which will include—

I. A brief description of the ruling machine (Rowland's) with a description of the process of making those parts of the machine which are of essential importance, and numerical estimates of the accuracy required in each.

II. An account of the method of testing the performance of the machine after it has been assembled, and a brief account of the method of testing a grating.

I. The Ruling Machine

§ (1) General Description of Rowland's Ruling Machine.[1]—The main frame (36) (*Fig.* 1) supports two sets of ways, at right angles to each other. The plate carriage (11) moves on one of these, the *ruling* or diamond carriage (5) on the other. The *feed screw* (12) rests in bearings supported by the main frame and is prevented from moving longitudinally by the *thrust screw* (14). As the feed screw is rotated by turning the *spacing wheel* (23) the *nut* (15) moves toward the latter, and by means of the *thrust collar* (20) its motion is communicated to the plate carriage. The ruling carriage ways consist of two pairs symmetrically placed, one pair on each side of the plate carriage, as may be seen from *Fig.* 1. The ruling carriage is moved backwards and forwards by means of a crank on the *main drive shaft* (48A), the *connecting rod* (49), and the *cross head* (52). The main drive shaft is rotated uniformly by means of a belt passing over the wheel (48), the source of power being

[1] The numbers in parenthesis refer to the parts as shown in *Figs.* 1 and 2. (These figures are reproduced by permission of the Johns Hopkins Press, Baltimore, U.S.A.)

a water motor running under constant head. One revolution of the drive shaft may be spoken of as a cycle, for during this period one groove or "line" is ruled on the plate. We will take as the beginning of a cycle the moment when the ruling carriage is at the end of its stroke nearest the main drive shaft. At this moment the cam (55) is lifting the lever (54) which, acting through the rods (35 and 34) and the lever (56) and rod (57), lifts the *ruling diamond* (2) off the plate; the cam (47) has also lifted the *pawl lever* (26) to its highest position. During the first half cycle, while the ruling carriage moves forward the full length of its stroke, the diamond is held above the plate, being lowered just when the carriage stops at the end of its stroke. At the beginning of the forward stroke the lever (40) is lowered a little, causing the pawl to engage with the ratchet on the spacing wheel, and immediately the two levers (26 and 40) are lowered together a fixed amount, which rotates the spacing wheel and the feed screw a small fraction of a revolution, thus causing the plate carriage to move forward a distance equal to the grating space. As soon as the spacing has been accomplished the lever (40) is raised a trifle, disengaging the pawl from the spacing wheel. During the return stroke, or second half cycle, the ruling diamond is in contact with the plate, and the groove or "line"[1] is ruled.

In ruling a 5-inch grating with 15,000 grooves per inch this cycle is repeated 75,000 times, usually at the rate of about 20 per minute, so the time required is some 63 hours. If the grating is to be of good quality, it is obviously necessary that during this time the temperature must remain very nearly constant; the amount the plate carriage is advanced during each cycle must be so nearly constant that the actual position of the last groove will agree with its ideal position within a very small fraction of the distance between two adjacent grooves; the ruling carriage and the *diamond holder* (2) must perform their function so well that all the grooves shall be straight and not displaced in any way by accidental movements of any of these parts; and last, but not least, the ruling point must not change its shape during the whole process.

When we bear in mind that the whole machine is made of parts which are not rigid, but elastic, and that in actual operation many of the essential parts are deformed, by forces due to friction, to an extent *which may amount to several grating spaces*, we have reason to be surprised, not that occasionally a grating is faulty, but that any grating is ever good. After doing the very best we can in the matter of construction and adjustments there is no assurance that this will be sufficient; there is no "factor of safety." Let us try to make this important point a little clearer. As we shall endeavour to show presently, it is quite possible to make and adjust all the essential parts of the machine so accurately that if they were absolutely rigid the grating ruled would be so nearly perfect that the most sensitive optical tests would fail to reveal any error. Since, however, these parts are

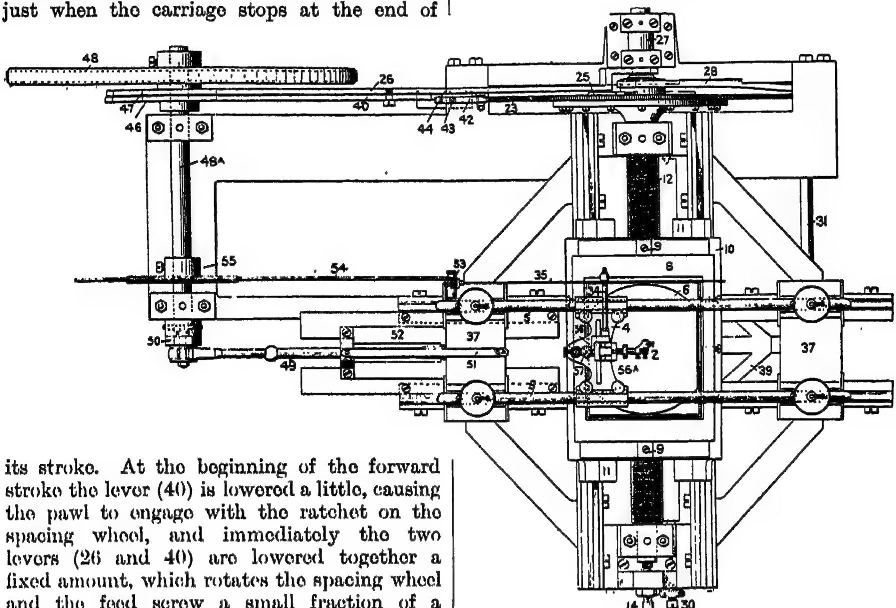


FIG. 1.

[1] In what follows we shall use the word groove, which is more nearly correct than the usual appellation, "line."

elastic, and consequently deformed by frictional forces, accuracy in the results demands that these forces and deformations *remain constant within rather narrow limits* throughout the entire process of ruling. This is the real difficulty in the manufacture of optical gratings, in comparison with which all others may be regarded as trivial.

Let us now enumerate the parts which we have spoken of as essential, beginning with those which, in general, present the greatest practical difficulties :

i. The ruling carriage with its driving connections.

ii. The connecting mechanism between the nut and plate carriage.

iii. The bearings and pivots of the feed screw.

iv. The thrust bearing for the feed screw.

v. The feed screw and its nut.

vi. The spacing wheel.

vii. The straight-edge guide for the wings of the nut.

Perhaps the reader will be surprised to find the feed screw placed fifth in this enumeration. There appears to be a general impression that the manufacture of a perfect screw is difficult, and that if it could be accomplished, the construction of very accurate measuring or ruling machines would be a simple matter. Nothing could be further from the truth. In fact, Professor Rowland himself clearly states that screws made by the process outlined by him,[1] on being tested failed to show any error as large as 1/100,000 inch, but that errors due to the mounting would certainly be encountered. The writer has made several screws by a process essentially the same as that described by Rowland, and has tested them by modern methods fully 10 times as sensitive as those employed by him, without ever discovering the slightest indication of any error. Errors due to the mounting are, however, less easy to avoid, and hence these have been placed 2nd, 3rd, and 4th in the enumeration.

The spacing wheel and straight-edge guide are placed last, because the error in the grating space produced by errors in these parts is only from 1/500 to 1/1000 of the error in the parts themselves, and hence nothing beyond ordinary mechanical accuracy is required.

The ruling carriage is properly placed first, and would, indeed, have been in a class entirely by itself except for a device found on Professor Rowland's machines which will be described more fully below. It is remarkable that Rowland never mentioned the ruling carriage in any of his writings; for its construction is such that even he must have had to give it considerable thought.

Professor Rowland did not deem it necessary to avoid entirely the periodic errors introduced by faulty mounting, but took care of these by an elaborate correcting mechanism working through the straight-edge guide. The writer had so much trouble with this correcting mechanism that he decided to do away with it altogether, preferring to eliminate the errors due to the mounting by careful attention to the items 2, 3, and 4 in the enumeration above. The advantage of this method is that when the machine is once in good adjustment it is likely to remain so at least for some years.

We will now discuss briefly the construction of the essential parts, remembering that periodic errors of spacing which in a half cycle add up to a quantity as large as 1/1,500,000 inch must be avoided, and that errors not periodic in nature must be kept correspondingly small.

§ (2) Construction of the Feed Screw.—A grade of well-annealed tool steel employed by manufacturers of taps and dies is used for the screw, and, if possible, also for the nut. The cutting is best done on a lathe having two tool rests [2] so that two tools, one on each side and 180° apart, may be used in cutting the threads. This automatically removes the greater part of the error in the head of the lathe. Let us suppose that the finished thread which is to be of the V-type is to have an angle of 52°. Three pairs of tools ("gooseneck" or spring type) should be prepared, the angles being 45°, 52°, and 60°. Cut to full depth of the finished thread with the 45° tools; then, use the 60° tools, cutting until the tops of the threads are sharp; finish with the 52° tools.[3] This being done, the pivots are turned to size, the taper to receive the spacing wheel is cut, and the screw is ready for grinding. It is a waste of time to take any elaborate precautions against introducing errors due to the lead-screw or gears of the lathe. A lathe is not, and never can be, a high precision machine.

Three nuts should be prepared, using for the purpose steel tubing whose finished internal diameter is about ⅜-inch larger than the outside diameter of the screw. One of the nuts should have a length equal to ¾ the length of the screw; another should be very short, say an inch, or two at most. The third nut should be just twice as long as the nut which is finally to be used on the screw; in fact, the latter is to be one of the halves into which the third nut is finally cut. The longer nut will be spoken of as the *grinding nut*, the very short one simply as the *short nut*, while the third one is usually designated the *testing nut*. Each is split in two by a longitudinal cut

[1] *Encyclopædia Britannica*, article "Screw."

[2] This device is due to Mr. L. E. Jewell, formerly of the Johns Hopkins University.

[3] This is the method employed by Mr. C. Jacomini of the Mount Wilson Observatory.

through its axis.[1] A brass cylinder previously machined to fit snugly inside the nut casings, having walls $\frac{1}{4}$ inch or a little more in thickness, is now cut into rings each about $\frac{1}{2}$ inch in length. Each of these is, in turn, cut into two parts by longitudinal cuts, and the halves are fastened securely in the casings of the grinding nut and the short nut, by screws; care being taken to leave about $\frac{1}{16}$-inch clearance between them. The testing nut is treated similarly, except that Babbitt's metal[2] is used instead of brass. This is best done by casting in place, no screws being required, since the molten metal will flow readily into the countersunk holes provided for the purpose.

The grinding nut is now bored to size and threaded, the two halves being clamped together with a spacer about $\frac{3}{16}$ inch thick between them. The short nut is treated in a similar manner, but it is better to postpone the threading of the testing nut until the process of grinding is nearly completed.

For grinding, the screw is supported between the centres of a lathe, and rotated at a speed of from 60 to 120 turns per minute. The grinding nut is balanced by a counterweight, and prevented from rotating by a simple lever held in the hand of the operator, who is thus enabled to regulate the friction by feeling, and who will also be warned immediately if any accident should occur. The nut is clamped on the screw by two or four rings held in place by pointed screws in contact with the two halves of the nut. To prevent slight rotations of the halves relative to one another, strips of corrugated spring steel or brass may be inserted in the slots between the nut casings. Emery or carborundum (washed 5 minutes) and oil is used as the grinding material. The process is carried out in air, and no extreme precautions against temperature changes in the room are necessary. One must, however, be careful not to heat the screw by too much friction or too great a speed. The nut is rotated 180° about the axis of the screw frequently; and once, or at most twice, a day it should be turned end for end; occasionally, only one of its halves should be turned end for end.

It is a good plan to wash both screw and nut at the end of a day's work for purposes of inspection. After two or three days' grinding a glance at the screw from a distance of a few feet will reveal all the errors introduced by the lathe. The appearance is so striking that it need not be described. Grinding should be continued for a short time after all these irregularities have disappeared, which will usually require from four to twelve weeks. The diameter of the screw should be frequently measured at several points, by a micrometer caliper, to be sure that it is everywhere the same, and also to see how fast it decreases under the grinding. It is usually found that the process reduces the external diameter by from 0·010 to 0·020 inch. The operator will soon be able to detect variations in the diameter of the order of 1/20,000 inch with certainty. When the first stage of the grinding has been completed, the short nut is fitted by grinding, which may require a day or so, care being taken to run it evenly throughout the full length of the screw. The parts are then thoroughly cleaned, and the short nut replaced, using only oil with no grinding material. A lever attached to the short nut is held by means of a spring balance, and the reading of the balance is noted as the nut is run from one end of the screw to the other. In this way a very accurate test for constancy of the diameter is obtained. If this is not constant it must, of course, be corrected by further use of the grinding nut.

This is followed by a few days' grinding with 30-minute-washed emery or carborundum, after which the short nut is again used to test the diameter.

The testing nut is now threaded and fitted to the screw, using the 30-minute-washed emery. If this requires more than two days the diameter should again be tested.

The screw is then polished, using oil and rouge, first with the grinding nut, and then with the testing nut. It is now finished, except that it is necessary to remove completely all traces of grinding material and rouge, which are likely to be imbedded in both screw and nut. This is accomplished by using the testing nut with oil only, washing thoroughly at the end of each day's work. *It is important that this process be continued a few days longer than the operator thinks it is necessary.*

§ (3) Testing the Feed Screw.—The testing nut is cut into two nuts of the same length, to each of which is added a steel wing about 7 inches long. The screw is mounted in its bearings on a suitable frame which carries a straight-edge against which the wings of the nuts may rest, to prevent them from turning when the screw is rotated. One plate of a Fabry and Perot type of interferometer is mounted on each nut, so that the line of sight through the interferometer shall be parallel to the axis of the screw and 4 or 5 inches vertically above it. A mercury lamp is a good source of light for use with the interferometer.

If in one rotation of the screw the interference rings remain stationary, the axis of

[1] The writer had to discard the four-part grinding nut recommended by Rowland. It is impossible to adjust it properly during the process of grinding.

[2] The use of Babbitt's metal instead of wood in various parts of the machine is due to Mr. L. E. Jewell.

the screw is straight. An easy calculation shows that this method will readily detect a curvature of the axis having a radius of 300 miles. If, however, the steel used in making the screw is of good quality and well annealed, and if the work of machining has been done in such a way that no strains have been introduced, no curvature of this amount will be found. It is interesting to bear in mind that the screw mounted in this way sags down in the middle under its own weight, so that the actual radius of curvature is anywhere from 2 to 10 or 20 miles; but as the screw is rotated, this radius always points upwards, so that it does not affect the interference rings. A clear mental picture of this cannot help but be of immense value to any one who has to deal with apparatus where great accuracy is a consideration.

By repeating the interferometer test with one of the nuts rotated 90° or 180° relative to the other, the operator may satisfy himself in regard to the absence of any periodic error; by separating the two nuts a considerable distance, and repeating the test, errors of run may be investigated. For this it is, of course, necessary to watch the interference rings while the nuts are moved a distance of several inches, if the test is to be sufficiently sensitive. No errors of any kind will be found except very close to the ends of the screw. The operator will, however, discover that it is no easy matter to put two well-fitting nuts on a screw without introducing anything between the threads except oil. If the fit is perfect (nothing but oil between them) pressure applied to *any* part of either nut will displace the interference rings, but on removing the pressure they will return exactly to their original position. If a particle of foreign matter has been introduced this will not be the case.

It should be noted that a screw made as described above will not necessarily have "perfect threads" as defined by the specifications of our national physical laboratories during the late war. Those specifications were intended to apply to parts which are to be interchangeable. The screw and nut described above form a unit; another nut, for example, cannot be used with the screw, without first being ground to a fit with all due care.

§ (4) Adjusting the Pivots of the Feed Screw.—One of the nuts is removed, and the other is moved near one end of the screw, but not so close as to engage the imperfect part of the thread. One interferometer plate is mounted on this nut so that the normal to its surface is vertical and fairly accurately at right angles to the axis of the screw. The other interferometer plate is supported from the frame and mounted above the first plate. On rotating the screw it will be found that the lower plate describes a small circle with its axis parallel to that of the screw. This shows that the lathe did not turn the pivot in such a way that its axis coincides exactly with that of the screw. The pivot must then be corrected locally, and be kept truly cylindrical by grinding in its own bearing, until the circle is quite small, say, until its diameter is of the order of 1/100,000 inch. It is not advisable to carry the correction further at this stage, for the other pivot is probably contributing something to the apparent error. The nut is moved near the other end, and that pivot is treated the same way, only that the correction is carried somewhat further, perhaps to 1/250,000 inch. Then, return to the first pivot, and so on until no error can be detected. The pivots are finally polished with rouge, and again tested as described above.

Three part bearings should be used with the screw pivots, making certain of contact at three points of each pivot, these points being 120° *apart.* The proper design and construction of such bearings present no serious difficulty.

§ (5) The Thrust Bearing.—One end of the screw, outside the pivot, is provided with a taper bearing for the spacing wheel. In the other end four holes are drilled and tapped for receiving the screws used to fasten the mounting of the thrust plate. It is advisable to finish this end of the screw optically flat, the surface being at right angles to the axis within about one or two seconds of arc. Various materials for the thrust plate have been tried, but none have been found as satisfactory as the reconstructed ruby, especially those of French make. A cylinder of this material $\frac{3}{16}$ inch to $\frac{1}{4}$ inch long, and from $\frac{1}{4}$ to $\frac{3}{8}$ inch in diameter. is finished optically flat and highly polished, at least on one face. No error as large as $\frac{1}{20}$ of a light wave should be allowed. The ruby is rigidly secured in a steel mounting, which is fastened to the end of the feed screw by four well-made screws as indicated above. The surface of the mounting in contact with the flat end of the feed screw must be worked until, when the two are in contact, the face of the ruby is also normal to the axis of the screw within one or two seconds of arc.

All of this requires some patience but is not especially difficult to accomplish. Suitable optical tests are, of course, needed to ensure the required accuracy in alignment, and this is one reason why the face of the ruby must be so flat. If it is not flat to $\frac{1}{20}$ of a wave-length, the diffraction disc observed at the focus of the lens or mirror used in the test will not be truly circular, thus making it difficult to observe a motion of $\frac{1}{10}$ of its diameter when the screw is rotated.

The end of the thrust screw (14) is a hardened steel spherical surface, highly polished.

having a radius of curvature of from 2 to 6 inches. It should make contact with the ruby at a point on the axis of the screw, but, thanks to the accuracy of alignment of the surface of the ruby, an error of 1/100 inch in the position of the contact can at most produce a periodic to-and-fro motion of the screw of something like 1/5,000,000 inch, which is too small to produce any harmful effect.

A thrust bearing of this kind, after continuous use for several years, will not show the faintest indication of a mark due to wear on either surface; the force acting across the contact may safely be as high as 40 lbs. No doubt a diamond would be as good or even better than a reconstructed ruby, but it would cost considerably more and would be more difficult to work. Agates, topazes, natural sapphires, etc., have been tried and found unsatisfactory.

§ (6) The Connecting Mechanism between the Nut and Plate Carriage.—The original design by Professor Rowland, described in his article on the screw,[1] and shown in *Fig.* 2 (20), is still followed. This consists of a ring or collar loosely supported by the plate carriage, and making definite contact with it at two points near the ends of a vertical diameter. The nut makes contact with the ring at two points near the ends of a horizontal diameter. In each case the contact is made by a hard flat surface against a softer rounded metallic lug. In one machine the two pairs of flat surfaces are mounted respectively on the plate carriage and the nut; on another machine, the flat surfaces are on the ring. It is important that the four contact points shall lie in one plane at right angles to the axis of the screw.

We will now try to give the reasons for placing this part second in the enumeration in § (1). The ring or collar is not difficult to make, and it is practically self-adjusting; but, located as it is at the most vital point in the whole machine, it is very likely to act as a trigger, bringing into action potential sources of error which otherwise would be held in abeyance.

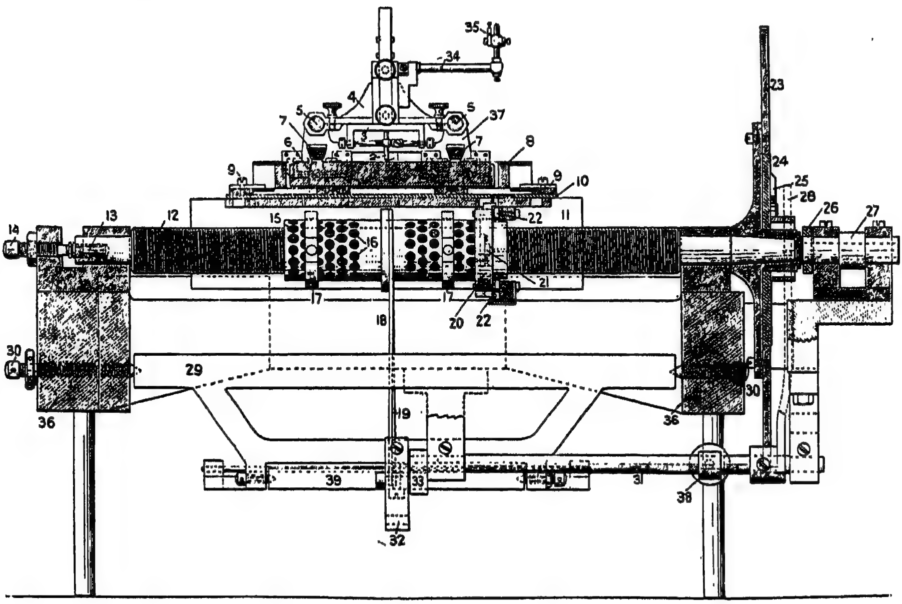


Fig. 2.

Let us examine a little more minutely the action of the ring. It is attached to the plate carriage in such a way that, when not in contact with the nut, it may move slightly in a direction parallel to the axis of the screw; it cannot move at right angles to this direction. It may rotate through a small angle about a vertical axis, but it cannot rotate about a horizontal axis, parallel to the screw. When the nut acts through it to move the plate carriage, its plates are brought into contact with the lugs of the latter, after which it has only one degree of freedom, namely a slight rotation about its vertical axis. Now since the axis of the screw can never agree *exactly* with that of its pivots, nor be exactly parallel to the plate carriage ways, it is evident that

[1] *Encyclopædia Britannica*, 9th edition.

as the screw rotates the nut describes a vertical circle of small radius, and in addition it moves very slowly in a direction at right angles to the ways of the plate carriage. The little circle is described once in each revolution of the screw, during which time the carriage moves forward, say 1000 times, 1/20,000 inch at a time. Suppose the diameter of the circle to be 1/200,000 inch, or its circumference, roughly, to be 1/60,000 inch. Assuming all the parts to be perfectly rigid, the lugs on the end of the nut, during each forward motion of the carriage, would slide along the plane surfaces on the ring a distance of 1/1000th of 1/60,000 inch or 1/60,000,000 of an inch. In the actual case, where the parts are not rigid but elastic, nothing like this can happen on account of the friction between the lugs and the plates. What really happens is probably this: no actual slip takes place until, owing to the relative displacements of the parts, the stress becomes sufficient to overcome the friction, when a comparatively large slip occurs, perhaps even "overshooting the mark." Remembering now that, during the normal operation of the machine, the screw is compressed a certain amount, and the nut, ring, and the parts of the plate carriage are all deformed as a result of the "forces of friction," and that *the successful performance of the machine depends upon these deformations remaining constant to a high degree of accuracy*, one can readily see what a catastrophe may easily result from the action just described.

The question naturally arises: "How accurately would the adjustments have to be made to ensure that no slippage whatever would occur during a revolution of the screw?" This the writer is unable to answer. A maladjustment to such an extent that a slippage would necessarily occur in each cycle would no doubt make matters much worse.

§ (7) THE RULING CARRIAGE.—Let us consider the task the ruling carriage has to perform in ruling say a 5-inch grating, having grooves 4 inches long. If the spacing is 20,000 per inch, it must move back and forth 100,000 times, or a distance of some 13 miles of actual travel, involving 200,000 stops and starts. It must move in one and the same straight line, deviating from this not more than 1/1,000,000 of an inch if the result is to be satisfactory. Moreover, the power that moves the carriage is derived from a main shaft which is called on to deliver power intermittently to several other parts at the same time.

Obviously, the friction coefficient must be made as small as possible, and the force required to start the carriage from rest should not differ greatly from that required to keep it in motion. Probably these conditions can not be satisfied with V-type ways. A better form of way is one having a flat top, with sides making an angle of 80° or more with the top. The weight of the carriage is thus mainly supported by the flat top; one of the sides serves to define the path of the carriage, and a spring acting on the opposite side ensures contact. The constant motion of the carriage causes an appreciable wear of the way surfaces. Measured in units of 1/1000 inch this wear is exceedingly slow; but it is relatively rapid when measured in the unit which concerns us here, namely, 1/1,000,000 inch. The wear of the top surfaces is of little or no consequence, as it merely displaces the carriage in a vertical direction, but that of the sides which define the path is of great importance. During the ruling of a single grating the sides change enough to make a considerable difference in the position and curvature of the first and last grooves. As the stroke is different for gratings of different size, one readily sees that after a year's work the defining sides of the ways will no longer be straight, but in reality quite irregular curves. In spite of these changes in the ways, and the wear of the bearing blocks of the carriage, Rowland's machines still rule grooves which are quite as perfect as those ruled at first when the ways were really straight. This result is due to following the device mentioned in § (1).

It will be recalled that the ruling carriage moves on two pairs of ways symmetrically placed, one pair on each side of the plate carriage, or we might describe it by saying that there is but one pair of ways with a section some 10 inches or more in length removed from the middle, in order to allow space for the plate carriage, its ways, and the screw (see *Fig.* 1 (37)). Let us call the pair farthest from the main shaft the front pair, the other the rear pair. *The defining sides of the front pair are on the left, those of the rear pair being on the right side, and the ruling point is at the centroid of the four contact points.* Hence if the two pairs of ways are straight initially, and the amount of wear is the same for both, the ruling point will always describe the *same* straight line. The wear will be the same provided the ways are made from the same material, and the pressure produced by the two sets of springs is the same.

The crank on the main shaft moves the cross-head (52), and a connection between this and the ruling carriage communicates the motion to the latter. The cross-head ways can, of course, never be made parallel to those of the ruling carriage, and hence this connection presents difficulties of the same nature as those discussed in § (6). A well-made universal coupling is employed and does fairly well, but it must be confessed that no entirely satisfactory solution has yet been found.

§ (8) The Spacing Wheel and the Straight Edge Guide.—As stated in § (1) the effect on the grating of errors in these parts is at most of the order of 1/500 of the error. Hence if these parts are correct to about 1/10,000 inch they will be satisfactory, and since this is little more than ordinary mechanical accuracy, we need not consider them further.

II. Testing the Machine and Gratings

§ (9) Method of Crossed Rulings.—Every one is familiar with the appearance of "watered silk," the pattern seen when looking through two moderately fine-meshed wire screens nearly parallel to each other, and separated by a small interval; or the pattern observed when viewing a screen and its reflection in a mirror immediately behind it. The principle illustrated by these familiar phenomena is used as follows to obtain a delicate test of the performance of a ruling machine.

A grating which may be $\frac{1}{8}$ inch wide is ruled on a well-polished plate. The plate carriage is moved back to the starting-point, and the plate holder is moved back an additional distance of from $\frac{1}{4}$ to $\frac{1}{2}$ the pitch of the feed screw, and given a rotation about a vertical axis of about one minute of arc, after which the plate is again ruled the same width as before. The second ruling crosses the first at the small angle mentioned, and on viewing the ruled surface in the light of one of its spectra a series of bright and dark bands will be observed, generally at right angles to the direction of the individual grooves. We speak of the dark bands as the locus of the intersections of the two sets of grooves, but in reality they merely run parallel to the loci. (A complete account of the phenomena observed with such a cross-ruling has never been given, and would apparently be quite difficult.) If the ruling diamond has been properly chosen and adjusted [1] the dark bands will be very narrow as compared to the distance between them, and hence will be well suited for testing the accuracy of spacing. If the spacing is perfect, the dark bands will be straight lines. normal to the grooves; errors of spacing will be indicated by departures from these conditions.

Choose rectangular axes, and take as the origin the intersection of the first groove of the first ruling with the first groove of the second ruling. (This point will not actually be on the plate, since the plate was displaced slightly before being rotated.) Let the X axis bisect the obtuse angle, the Y axis the acute between the two grooves.

[1] Ordinarily a diamond which will rule first-rate gratings will give only a poor cross-ruling pattern, the bands being so faint and diffuse that they are difficult to see.

(i.) *Ideal Spacing.*—Let the intersections with the X axis be given by

$$x_n = na(n=0, 1, 2, 3, \ldots), \qquad (1)$$

a, being the grating space, very nearly.

The equations of the two sets are

$$y_{1,n} = mx_n - mna, \qquad (2)$$

$$y_{2,n} = -mx_n + mna. \qquad (3)$$

We seek the intersection of any groove n of the first set with the groove $n+2p$ ($2p$, any integer) of the second set. The locus of these intersections is evidently

$$y = pma = \text{constant}; \qquad (4)$$

m is the tangent of the angle between any groove of the first set and the x axis, *i.e.* $m = \tan 89° 59' 30'' = 7000$ nearly. We will speak of $2p$ as the "order" of the intersection. Thus, for $p=0$, or the zero order, the locus is the x axis. For $p=\frac{1}{2}$ or the first order, $y=ma/2$, a straight line parallel to the x axis and at a distance $ma/2$ from it. Taking $a=1/15,000$ inch, which is the spacing usually employed in recent years, $ma/2$ is a little less than $\frac{1}{4}$ of an inch, which is the distance between the "orders."

(ii.) *Simple Periodic Error.*—Let

$$x_n = na + b \sin \frac{2\pi n}{N}. \qquad (5)$$

The locus of intersections of order $2p$ now becomes

$$y = mpa + mb \sin \frac{2\pi p}{N} \cos \frac{2\pi(n+p)}{N}. \qquad (6)$$

N is the number of grooves ruled in a period, and as the fundamental period will always correspond to one revolution of the screw, N is simply the number of teeth on the rim of the spacing wheel.

Again the zero order ($p=0$) reduces to the x axis; also, $2p=N$, $2N$, etc., give straight lines parallel to the x axis. Other values of p give cosine curves, the amplitude of which is a maximum for $2p=N/2$, $3N/2$, etc.

The amplitude of the curve whose order is $2p$ is $mb \sin 2\pi p/N$. If this equals $ma/4$ we speak of it as "unit amplitude." Such a choice of unit is necessary, as the actual amplitude depends upon m, which changes very rapidly with the angle, and, even when great care is taken to keep the angle constant, will vary considerably in different cross-rulings. The value of b which gives unit amplitude in the $2p$ order is evidently

$$b_0 = \frac{a}{4 \sin(2\pi p/N)}.$$

Let the actual amplitude as measured in

terms of the unit defined above be A (where in general $A<1$). Then our error is

$$b=\frac{aA}{4\sin(2\pi p/N)}=\frac{aA}{4\sin(\pi S/N)}, \quad . \quad (7)$$

S being the "order."

As a numerical example, let us take $A=\frac{1}{25}$, which under ordinary conditions is fairly easy to observe. If $a=1/15{,}000$ inch, $S=N/2$, we have $b=1/60{,}000\times 1/25=1/1{,}500{,}000$ inch as the maximum periodic displacement of the plate away from the position it should have during ruling.

We may also compute the value of the maximum and minimum grating space in such a ruling. Referring to formula (5) and taking $N=750$, we have

Max. spacing $=1/15{,}000+1/190{,}000{,}000$ in.

Min. spacing $=1/15{,}000-1/190{,}000{,}000$ in.

A variation of the spacing as large as this, if it were not periodic, would, of course, destroy the value of the grating as an optical instrument. (This should be borne in mind in connection with the discussion in § (6).)

It is well known that the effect of a simple periodic error is to produce ghosts of bright lines [1] in a spectrum. If the error is not too large and is simply periodic in character, there will be two ghosts of each line, symmetrically placed with respect to it, at a distance which varies inversely with N. It is a simple matter to calculate the intensity of such ghosts when b is known, provided it is small compared to a. The derivation of the formula is, however, rather lengthy,[2] and therefore only the formula itself will be given here. Taking the intensity of the spectrum line as unity, the intensity of the ghost is given by

$$I_g=\frac{M^2}{4}\left(\frac{2\pi b}{a}\right)^2, \quad . \quad . \quad (8)$$

M being the order of the spectrum observed. If $b/a=1/100$, as in the example given above, we have

$M=$	1	2	3	4	5
$I_g=$	0·001	0·004	0·009	0·016	0·025,

or in the 5th spectrum the ghosts have an intensity equal to $\frac{1}{40}$ of that of the line itself. In a number of the gratings ruled since 1911 the ghosts are rather weaker than this.

(iii.) *Linear "Error of Run."*—If the spacing varies linearly across the grating, it is commonly spoken of as showing an "error of run." The law of spacing is

$$x_n=na+n^2c, \quad . \quad . \quad . \quad (9)$$

where $2c$ is the constant difference between any grating space and the preceding one. The equation of the cross-ruling band of order $2p$ is

$$y=mpa+2mcp(n+p)\,; \quad . \quad . \quad (10)$$

a straight line whose slope is $2mcp$.

In order that the test for this type of error shall be sensitive, it is necessary to take p quite large, and also to rule quite a wide space, so that the length of the band may be great enough to allow an accurate determination of its slope.

As a numerical example, let us consider a ruled space 4 inches wide, 15,000 grooves per inch, and let the plate be displaced 4 inches between the first and second rulings. Let y_1 and y_2 be the distances from the x axis to the beginning and ending of the band of order $2p$. We have

$$y_2-y_1=2mcp(n_2-n_1). \quad . \quad . \quad (11)$$

In this case $p=3\times 10^4$; $n_2-n_1=6\times 10^4$; m may be taken as 7×10^3. Suppose $y_2-y_1=\frac{1}{25}$ inch, which should be measurable. We have

$$c=1{\cdot}6\times 10^{-15} \text{ inch.}$$

The departure of the last groove from its ideal position on a 4-inch grating having this error is $n^2c=5{\cdot}8\times 10^{-6}$ inches, or about $\frac{1}{4}$ of a light wave.

The two illustrations just given will suffice to show that in cross-ruling we have a powerful method for studying the performance of the machine. Other cases might be discussed, such as variation in spacing along the length of the grooves and curvature of the grooves, but the reader will have no difficulty in seeing how these cases are investigated.

(iv.) *Accidental Errors and Diffused Light.*—Accidental errors are due to two principal causes:

(*a*) That discussed in § (6).

(*b*) Slight errors in the motion of the ruling carriage, and those due to faulty adjustment of the diamond-holder (2, 3).

If (*a*) alone is present, the cross-ruling pattern will appear somewhat like *a* in *Fig.* 3. The band is clean and narrow but shows a large number of breaks, indicating that, superposed upon the regular, even displacements of the carriage, there are sudden shifts, which occasionally may amount to as much as $\frac{1}{10}$ of a grating space, but usually are very much smaller than this. If (*b*) alone is present, the pattern is very much as shown in *b*, *Fig.* 3. The pattern is not clean, but may be best described by saying it is "hairy." In general, both causes are present, so that an

[1] If the error is very large, ghosts may also appear in absorption spectra such as the solar spectrum, but such cases probably do not occur with any grating actually used in spectroscopic work. In ruling gratings one often finds them.

[2] It is easiest accomplished by using the so-called "vector method" of treating diffraction phenomena.

is in the *same* direction on both sides of the normal, which is not so when the error is in the ruling. (In the preceding discussion it has been tacitly assumed that the angle of incidence is zero, merely for simplicity. When the angle of incidence is different from zero, the different cases are affected somewhat differently, making it still easier to separate the causes of error.)

If the Foucault test is satisfactory, and periodic errors have been determined and found sufficiently small, the grating is satisfactory, provided it is also reasonably free from diffused light. The determination of the amount of diffused light, and also of the brightness of the different orders of spectra, do not require any comment.

§ (11) The Selection and Adjustment of the Ruling Diamond.—Natural crystal edges, free from little breaks or other imperfections, are used in ruling gratings. Small white diamonds, each weighing about $\frac{1}{30}$ carat, are securely mounted, each in the end of a short steel rod $\frac{1}{8}$ inch in diameter. The diamond-holder permits rotations about three mutually perpendicular axes in order that a given edge of the crystal may be brought into contact with the plate, and may also be adjusted parallel to the ruling carriage ways.

The diamond-holder can be removed from the machine and mounted under a microscope, where the preliminary adjustments are actually made, after which it is replaced in the proper position on the machine. The correct pressure of the diamond on the plate must be found by trial for each ruling edge, and is regulated by a counterweight.

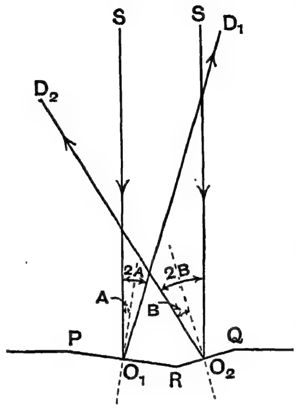
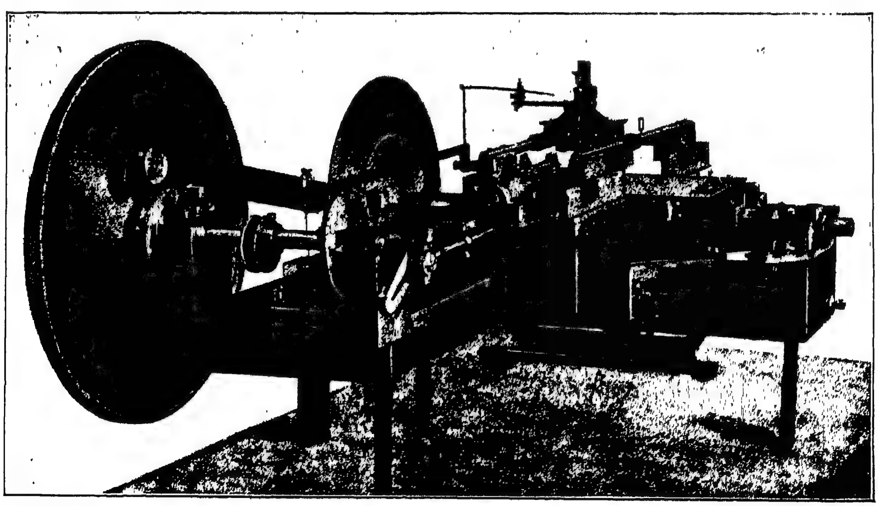


Fig. 4.

The ideal groove is V-shaped, and the actual groove approaches this form very closely if the adjustments are good. It remains to determine the angle of the groove, to give it the right width, and to orientate it properly for the kind of ruling desired.

Let a single groove be ruled on a well-polished plate. Suppose the normals to the sides of the groove make the angles A and B with the normal to the surface. If light be incident normally on the surface, the light diffracted by the groove will show two maxima, viz. in the directions making angles 2A and 2B with the incident beam. The diagram, *Fig.* 4, which represents a section normal both to the surface and to the groove, will make this clear. PR and QR are the sides of the groove; O_1D_1 and O_2D_2 are the directions in which the maxima of the diffracted light will be found. Since PR and QR are very narrow, having in general a width of from $1\frac{1}{2}$ to $2\frac{1}{2}$ wave-lengths of light, the angular width of the diffracted beams will


Rowland's Ruling Machine.

be considerable, and, in the figure, O_2D_2 will be wider than O_1D_1. It is, however, quite possible to measure approximately this width, that is, the angular separation of the first diffraction minima belonging to O_1D_1 and to O_2D_2, and this being done, the distance PQ becomes known. It should be as nearly as possible the same as the grating space, and may be adjusted to this value by varying the weight on the diamond. If more than two maxima are found, the groove is not of the simple V-type, and should in general not be used.

As it is desirable to have some particular spectrum as bright as possible, the diamond should be tilted until the direction of O_1D_1, for example, falls in that spectrum. The direction of O_2D_2 will in general be that of some higher order spectrum on the other side, but it cannot be controlled independently except by changing the grating space. The lower order spectrum which is bright at zero angle of incidence will also be bright for other values of this angle,[1] at least within reasonable limits.

This discussion is far from complete, but it is hoped that enough has been said to show that the finding of a suitable ruling point is not altogether a matter of chance. J. A. A.

DIFFRACTION GRATINGS, THEORY

I. Theory of Gratings

§ (1) Introductory.—A Diffraction Grating consists, in effect, of a number of parallel and equidistant slits in an opaque screen. The slits are usually very narrow and very numerous, but this is not invariably the case. If a parallel beam of light is incident on such an arrangement, some of it passes through the slits while some is stopped by the opaque intervals. If the transmitted light is brought to a focus by a telescope or camera lens, it is found that spectra are produced in certain directions. Such spectra are usually described somewhat vaguely under the comprehensive title "Diffraction Spectra." Physically there are three quite distinct processes contributory to the production of the phenomena observed: these are:

(*a*) *Diffraction* by the individual slits of the grating, converting them, in effect, into a series of similar light sources emitting light over a considerable range of angle on either side of the direction of incidence. The wavelets emanating from these sources start, however, with definite phase relationships; this renders possible the next stage, viz.—

(*b*) *Interference* between the wavelets from different slits, which results in their completely annulling one another except in certain directions.

[1] See *Astrophysical Journal*, 1911, xxxiii. 350.

(*c*) *Diffraction* by the aperture of the observing system. This gives finite breadth to the spectrum lines, and introduces a series of secondary maxima between them.

Needless to say these phenomena cannot be discriminated in the final result. Thus we do not observe any diffracted light resulting from process (*a*) except in those directions determined by process (*b*). Moreover it is impossible to observe the resultant of these two processes except under the limitations of total aperture set by the dimensions of the observing system, which introduces the further modifications due to process (*c*).[2] We cannot therefore dissociate the contributory processes in practice. This tempts us, in developing the theory of the grating, to put all the ingredients into our mathematical mill at once, and after mixing them thoroughly, obtain final equations from which we may extract all the essential information. This is the method followed in the classical treatments of the subject which are found with slight modifications in the majority of text-books. While of great mathematical elegance, such a treatment almost invariably fails to convey to the student any clear conception of why the various results which the equations indicate, and which he can see in practice, do actually happen. In order to understand these it is desirable to keep quite separate in one's mind the processes (*a*), (*b*), and (*c*). This desideratum is partly fulfilled in the vector method originally used by Kimball and subsequently employed by Wood, Sparrow, and others. In this treatment process (*a*), which merely supplies the light in the necessary directions, is dissociated from the subsequent spectrum formation. The analysis, however, automatically obtains the joint effects of (*b*) and (*c*) and is, in consequence, misleading as to the origin of secondary spectra and of the finiteness of resolving power. The method is of great use and beauty, nevertheless, and further reference will be made to it later; but the various properties of grating spectra will be deduced in the following paragraphs by a method which it is hoped will render their origin as clear as possible by treating each of the contributory processes separately.

§ (2) Diffraction by a Rectangular Aperture.—Although the starting point for any investigation of gratings, the theory of

[2] In any spectroscope the observing telescope is usually of sufficient aperture to take all the light from the prism or grating. In practice therefore the length of the grating and its obliquity determine the aperture of the system, but if in any case the telescope is of smaller aperture, in the plane of dispersion, than that required to take in light from the whole grating, it is the actual aperture which determines the results of process (*c*). For simplicity we shall assume in what follows that the aperture, whether due to the limited size of the grating or of the telescope objective, is bounded by straight sides parallel to the slits.

diffraction is outside the scope of this article. We must simply quote the result in the case of a rectangular aperture, as developed in any of the standard text-books.[1] Let A, *Fig.* 1 (*a*), be a point source of light situated at an infinite distance from a screen, B, in which there is a rectangular aperture *s*, with sides parallel and perpendicular to the plane of the paper. The light which passes through *s* does not pass straight on in the direction *s* A′, but spreads out in the space to the right of B, and would produce illumination over an area, more or less extensive, of a screen C. In what follows we need only concern ourselves with diffraction in the plane of the paper. It is also convenient in this case to substitute a line source, perpendicular to the paper, instead of a point source. Let the width of the aperture in the plane of the paper be a, then the light is spread out in this plane to an extent depending on a, the intensity in a direction making an angle θ with *s* A′ being proportional to the quantity

$$a^2 \frac{\sin^2 \{(\pi a/\lambda) \sin \theta\}}{\{(\pi a/\lambda) \sin \theta\}^2}, \quad . \quad . \quad (1)$$

where λ is the wave-length of the light.

In the direction *s* A′ $\theta = 0$, and the limiting value of the expression is a^2. When $(\pi a/\lambda) \sin \theta = p\pi$, p having the values 1, 2, 3, . . . etc., the intensity is zero, while it has maximum values at intermediate angles. If the light is received in a telescope focussed on infinity, so that all rays having the same value of θ are brought to the same focus, a series of bands will be seen corresponding to these alternate maxima and minima.

It is obvious from the form of expression (1) that the intensity curve of the lateral maxima is simply a sine squared curve of which the ordinates are divided by a quantity proportional to the square of the distance from the central band (measured in terms of $\sin \theta$). The proportional variation of this quantity is steep at first, with the result that the maximum ordinates of the maxima near the central band are displaced from the point half-way between the minima, being slightly nearer the central band. Their actual position corresponds to $\tan \{(\pi a/\lambda) \sin \theta\} = (\pi a/\lambda) \sin \theta$ instead of to $\tan \{(\pi a/\lambda) \sin \theta\} = \infty$. This follows at once from differentiation of expression (1). These values do not, however, differ greatly, and to a close approximation it is convenient to take the central maxima as occurring half-way between the minima, *i.e.* at the points when $(\pi a/\lambda) \sin \theta = (2p+1)\pi/2$,[2] beginning with $p = 1$. The ordinates at these points are given by

$$I \propto \frac{a^2}{\{(2p+1)(\pi/2)\}^2} = \frac{4I_0}{\pi^2} \frac{1}{(2p+1)^2},$$

where I_0 is the intensity at the centre of the central band. From this we find that the approximate maximum intensity in the first lateral band is 1/22·3 that of the central band; of the second 1/62; of the third 1/121, etc. The true values will be slightly greater than this, and can be calculated if required from the more accurate positions. The intensity curve is shown in *Fig.* 1 (*b*).

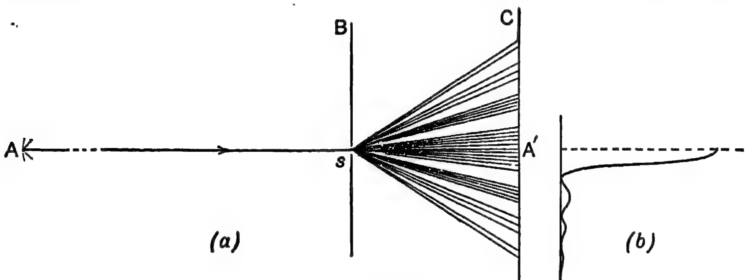


FIG. 1.

The positions of the lateral maxima depend on the wave-length, so if the light employed is not monochromatic, the bands corresponding to different colours will be in different positions. These bands were termed by Fraunhofer "*Spectra of the First Class.*"

If the width a of the aperture is very small the width of the central band, which extends to $\theta = \pm \sin^{-1}(\lambda/a)$, will be quite large. In fact if a is comparable with λ the central band will extend very nearly over the whole angular space on either side of the incident beam. Such a fine slit therefore acts practically as a line source of light as far as direc-

[1] *E.g.* Preston, *The Theory of Light*, chapter ix.; see also "Light, Diffraction of."

[2] The actual departures from this may be calculated easily by giving $\pi a/\lambda \sin \theta$ its approximate value corresponding to the middle of the band, and using this to obtain a closer value from a table of tangents. Thus for the first band $\pi a/\lambda \sin \theta$ is in the immediate neighbourhood of $3\pi/2 = 4{\cdot}7$; the true value will be very nearly $\tan^{-1} 4{\cdot}7 = 3\pi/2 - 12°$. Similarly the second maxima occurs at $5\pi/2 - 7°$, and the third at $7\pi/2 - 5°$. Since the width of a band corresponds to 360°, the first maximum is displaced by 1/30 of the width of a band, the second by 1/51, and the third by 1/72. As the order of the bands increases the distortion quickly becomes negligible.

tions in a plane perpendicular to its length are concerned.

§ (3) GRATINGS.—If we have a number of similar slits situated side by side the diffraction bands from each of them coincide in direction, and therefore in position if observed in a telescope. If the slits were illuminated by independent sources of light, the observer would see a system of diffraction bands exactly similar to those from a single slit, but more intense on account of the superposition of the systems from the several slits. In a grating, however, all the slits are illuminated by the same source. The diffracted waves emanating from them have therefore a constant phase relationship and so are capable of mutual interference. Since we are now regarding the grating elements as separate sources it is convenient for the time being to assume uniform radiation in all directions. The effects of the actual variation which takes place in accordance with the results of the previous paragraph can be investigated later.

Let the dots, *Fig.* 2, represent the equidistant slits of a grating on which a plane

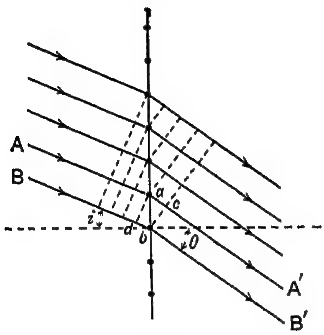


FIG. 2.

wave of perfectly monochromatic light is incident at an inclination i to the normal. In order that we may at this stage avoid the complications introduced by diffraction by a restricted aperture we must regard the grating as extending to infinity in both directions. Clearly any ordinary grating may be regarded as an infinitely long grating viewed through a restricted aperture.

Let Aa and Bb be rays incident on corresponding points of two adjacent slits, while aA′ and bB′ are rays diffracted in the direction θ. The disturbance from a lags behind that from b by a path difference $ac - bd = \Delta(\sin\theta - \sin i)$. In this expression $\Delta = ab =$ the distance between corresponding points of adjacent slits. The path difference between any slit and the rth from it is clearly $r\Delta(\sin\theta - \sin i)$, so that if $\Delta(\sin\theta - \sin i)$ is a whole number of wave-lengths, say $m\lambda$, the path difference between the disturbances from any two slits whatever is also a whole number of wave-lengths and all the disturbances are in the same phase. Hence if these are combined by a telescope, they will reinforce each other and the field will be bright at the position corresponding to this direction.

For directions differing slightly from the values of θ given by the above relation there is a phase difference, $\delta\lambda$, between the disturbances from adjacent slits. It is clear that however small $\delta\lambda$ may be, the resulting illumination due to the whole of the infinitely long grating is zero, because, since the phase step accumulates from one slit to the next, whatever the phase of one slit there can be found another slit for which the phase differs by $\lambda/2$. These two slits nullify each other. Similarly for all other slits. The grating therefore concentrates the light in a series of bright lines of infinitesimal width at infinity, in such directions that

$$\Delta(\sin\theta - \sin i) = m\lambda, \quad . \quad . \quad (2)$$

with complete darkness in the intervening intervals. Those lines for which $m = 1, 2, 3,$ . . . etc., are said to be of the 1st, 2nd, and 3rd, . . . etc., order. Equation (2) is referred to as the grating law. Those lines, of which the position depends on the wave-length, were termed by Fraunhofer "*Spectra of the Second Class.*" They are the true spectra of the grating *qua grating*, and we observe that they are not a diffraction phenomenon at all but are due purely to interference between the radiations from a series of equally spaced similar sources of light. Their position depends simply on the separation, Δ, of the sources, and is unaffected by their dimensions, precisely as in the case of interference fringes produced by, say, Fresnel's Mirrors or Biprism,[1] and it is no more legitimate to attribute them to diffraction than to attribute bi-prism fringes to refraction. The part played by diffraction in the action of the diffraction grating is, as we saw in § (2), the subsidiary one of converting the individual slits into sources of light radiating over a wide angle within which the interference phenomena may take place.

§ (4) EFFECTS OF FINITE APERTURE.—Such a simple instrument as the ideal grating of infinite length is naturally unrealisable. The finite size of any practical grating introduces very profound modifications in the observable phenomena. Nevertheless such modifications cannot properly be classed as grating action because they are precisely those modifications, and nothing more, which any optical system introduces in the image of a point or line source at infinity. The proper way to regard this part

[1] See article on "Interference."

of our subject is to consider the grating alone as supplying a spectrum consisting of infinitely thin lines at infinity, and then to consider the effect of examining these in an optical system of aperture determined by the size of the actual grating, or of the telescope, whichever is least.

If A is the effective aperture measured perpendicular to the axis of the telescope, the image of a true line of light consists (§ (2)) of a series of diffraction bands, the intensity and location of which are given by the formula

$$I \propto A^2 \frac{\sin^2\{(\pi A/\lambda)\sin\phi\}}{\{(\pi A/\lambda)\sin\phi\}^2},$$

where ϕ is measured from the position of the geometrical image. The points of zero illumination occur when $(\pi A/\lambda)\sin\phi = p\pi$, where $p = 1, 2, 3, \ldots$ etc.

The angular semi-width of the central band and the width of the lateral bands in its vicinity is λ/A.

Each spectrum line is therefore represented by a diffraction band of width $2\lambda/A$ flanked by a series of fainter bands of half this width. The maximum intensity in the pth band is $I_0(4/\pi^2) \cdot 1/(2p+1)^2$ where I_0 is the maximum intensity in the central band. These are the so-called secondary maxima. We see that they are in no way uniquely connected with gratings. The same type of central image with the same maxima in attendance is obtained with a prism spectroscope, or, for that matter, with the direct image of the slit, provided the same aperture be used. The only properties of secondary maxima peculiar to gratings become manifest when the number of grating elements is small, and are due to the overlapping of the secondary maxima associated with the primaries of different orders.

A primary occurs, as we saw in the last paragraph, whenever the phase difference between disturbances from corresponding points of adjacent grating elements has the value zero. A secondary minimum occurs whenever the phase difference between opposite edges of the aperture has the value zero, that is to say, when the phase difference between the *first* and *last* of the effective grating elements is zero. Clearly, therefore, if there are n elements, a primary of any order will coincide with the nth secondary minimum associated with the primary of the next order, and *vice versa*.

There are therefore $(n-1)$ minima between the two primaries, and consequently $(n-2)$ secondary maxima. With ordinary gratings the separation of the orders is so large compared with the distance between the secondaries that those associated with one primary are quite outside the influence of those associated with the next. The intensity of the secondaries diminishes so rapidly with p that the great bulk of the space between the spectra is quite dark. In fact the aperture has to be made quite small or a very high magnification has to be used to see any secondary maxima at all. If the total number of grating elements is small, however, the primaries of successive orders may only be separated by a few secondary maxima, and those associated with the different primaries will reinforce each other.

If we remember that there are $(n-2)$ maxima between each primary, and that the latter occupy the width of two of the former, it is easy to see that the pth secondary maximum associated with the primary of the mth order coincides with the $(n-1-p)$th, $(2n-1-p)$th, $(3n-1-p)$th . . . etc. secondaries of the $(m+1)$th, $(m+2)$th, $(m+3)$th . . . etc. orders, and with the $(n+p)$th, $(2n+p)$th, $(3n+p)$th . . . etc. secondaries of the $(m-1)$th, $(m-2)$th, $(m-3)$th . . . etc. orders.

The total intensity of this secondary is therefore

$$\frac{4I_0}{\pi^2}\left\{\frac{1}{(2p+1)^2}+\frac{1}{(2p-2n+1)^2}+\frac{1}{(2p-4n+1)^2}+\frac{1}{(2p-6n+1)^2}+\ldots+\frac{1}{(2p+2n+1)^2}+\frac{1}{(2p+4n+1)^2}+\frac{1}{(2p+6n+1)^2}+\ldots\right\}. \quad (3)$$

From this we can evaluate the intensities of the secondary maxima. As would be expected, except when n is very small, all but the first term in the bracket is negligible for secondaries in the neighbourhood of the mth primary, and all but the second term for those near the $(m+1)$th, and so on. Even for $n=4$ it is only necessary to consider the effect of the two adjacent orders on either side of the order in question.

The most intense secondary is the one nearest the primary. Its maximum intensity relative to that of the primary is about 1/16 for 4 elements, 1/21 for 8, 1/22 for an infinite number of elements. Thus there is little difference between the relative intensities of the strongest secondary whether a grating has an infinite number of elements or only a few. This is sometimes regarded as a surprising result, but when we realise that the secondary maxima are merely the ordinary diffraction bands due to the aperture of the system, we see why they are unaffected by the nature of the grating, except when the successive orders of spectra are so close that the relatively strong secondaries associated with the different images overlap. Moreover, this is a pure superposition effect; if we place a number of equidistant slits at the end of a collimator and adjust the aperture of an observing telescope to such a value that the angular width, λ/A, of the diffraction bands bordering each image is exactly $1/q$th of the angular separation of the images, we obtain precisely the appearance of several orders of spectra produced by a grating of q elements, with the appropriate secondary maxima between, without the presence of any grating whatever.

Expression (3) is symmetrical for values of p on either side of $p=(n-1)/2$, which represents the point half-way between the primaries; the reason of this symmetry is obvious from what has gone before.

§ (5) Vector Method.—Expression (3), obtained by adding the effects of the superposed diffraction bands which border the primaries of different orders, represents by its structure the actual method of formation of the secondary maxima. It is not so convenient for calculation purposes as that which is directly obtained by the vector method, in which the net result of interference and diffraction by the aperture is deduced in one process. We shall briefly outline this method, which is of great utility in many of its applications. Let there be n elements in the grating, and let the amplitude of the disturbance from each element be k. Let the phase difference between the disturbances from corresponding points of adjacent elements be $\delta\lambda$. It is convenient to express this in angular measure ϕ, where $\phi=(2\pi/\lambda)\delta\lambda$. The amplitude of the resultant disturbance is represented by the gap between the extremities of the first and last sides of a vector polygon composed of n sides of length k (on some suitable scale), each inclined to the one preceding it by the angle ϕ. Since the sides and angles are equal, the polygon lies on a circle, of radius such that k is the chord of ϕ. The size of the circle varies with ϕ, but at any stage the gap between the free ends of the first and last sides is

$$k\times\frac{\text{chord } n\phi}{\text{chord } \phi},$$

which is the amplitude of the resultant disturbance. The intensity is proportional to the square of the amplitude and is therefore

$$Bk^2\frac{\text{chord}^2\, n\phi}{\text{chord}^2\, \phi}=Bk^2\frac{\sin^2(n\phi/2)}{\sin^2(\phi/2)},$$

where B is a constant.

In those directions for which ϕ is zero, *i.e.* when the phase step between adjacent elements of the grating is an integral number of wave-lengths, the value of this is n^2Bk^2, which is the intensity at the centre of the principal maxima. Denoting this by I_0 we see that the intensity in intermediate directions is

$$I=I_0\frac{\sin^2(n\phi/2)}{n^2\sin^2(\phi/2)}. \quad . \quad . \quad (4)$$

This is zero when $n\phi=p\pi$, where $p=1, 2, 3, \ldots n-1$.

The $(n-2)$ secondary maxima occur between these minima, where $n\phi=(2p+1)\pi/2$, their intensities being given by

$$I_p=I_0\frac{1}{n^2\sin^2(2p+1)\pi/2n}. \quad . \quad (5)$$

The secondary maxima were termed by Fraunhofer "*Spectra of the Third Class.*"

§ (6) Dispersion and Resolution.—The rate at which θ, in equation (2), varies with wave-length is a measure of the spreading out of the spectrum, and is termed the *dispersion.* By differentiation of equation (2),

$$\frac{d\theta}{d\lambda}=\frac{m}{\Delta\cos\theta}=\frac{\sin\theta-\sin i}{\lambda\cos\theta}, \quad . \quad . \quad . \quad (6)$$

from which we see that for a given angle of incidence the dispersion in the neighbourhood of wave-length λ depends simply on the direction, θ, in which the spectrum is formed. It is independent of the properties of the grating, a large or small grating space Δ being compensated by the higher or lower order of the spectrum formed in the direction required.

It is not the dispersion which is of greatest importance in determining the utility of a spectroscope, but the *resolving power,* that is the power of producing two distinguishable images for two monochromatic radiations of slightly different wave-lengths. This power is usually measured by the ratio $R=\lambda/\delta\lambda$, where $\delta\lambda$ is the smallest difference which can be resolved at wave-length λ.

We saw in § (3) that the grating alone, in the absence of diffraction at the bounding aperture, would give infinitely fine lines. Its resolving power would therefore be infinite, since, however close two lines might be, they would still be perfectly distinct. However, the lines can never be observed except as diffraction bands, and if two of these are too close together they will merge into a single maximum of illumination and will not be seen as separate bands. Following Lord Rayleigh (7), it is always conventionally assumed in calculating the resolving power of optical instruments that two diffraction images will be resolved when the centre of one falls on the first minimum of the other, as shown in *Fig.* 3. The angular

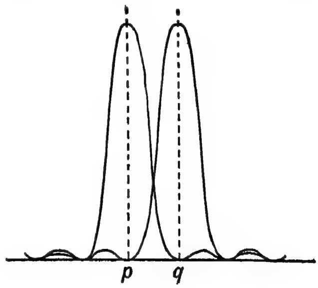


Fig. 3.

value of pq is $\delta\theta$, $=\lambda/A$, A being the effective aperture. The difference in wave-length comprised in this angle is

$$\delta\lambda=\delta\theta\cdot\frac{d\lambda}{d\theta}$$

$$=\frac{\lambda}{A}\cdot\frac{\Delta\cos\theta}{m}. \quad \text{(From 6.)}$$

L, the total length of the grating which is utilised, is equal to $A/\cos\theta$,

$$\therefore \delta\lambda = \frac{\lambda}{L}\cdot\frac{\Delta}{m}.$$

Whence $$R = \frac{\lambda}{\delta\lambda} = m\frac{L}{\Delta},$$

$$= mn. \qquad (7)$$

Thus the resolving power is equal to the product of the total number of grating elements and the order of the spectrum; but while this is a convenient form in which to remember the result, the apparent dependence on n is illusory. In any given direction Δ/m is constant, and R depends simply on the total length, or, more strictly, on A. It is only when varying the number of elements *while keeping the grating space constant* that altering n alters the resolving power, since in this case an increase in n involves an increase in L and *vice versa*, but for a given available aperture the resolving power does not depend on how many or how few grating elements we introduce.

It facilitates the understanding of spectroscopic resolution if we always separate, as in the above treatment, the two factors on which it depends. There is first the resolving power *in angle*, which is purely a function of the aperture of the observing system, the minimum angle for resolution being λ/A (assuming a rectangular aperture). The difference in wave-length corresponding to this angle depends on the dispersion of the spectrum whether produced by interference or prismatic refraction. The usual methods of deducing resolving power tend to conceal that its finiteness is purely a question of aperture, and lead the student to suppose that each type of spectroscope is a law unto itself in the matter of resolving power, which appears to depend on numbers of rulings, or length of prism base, etc., whereas these are merely quantities which happen to be conveniently related to the aperture of the system for formula purposes.

Reverting to the special case of gratings, suppose λ' is the wave-length in the mth order which coincides with λ in the $(m+1)$th order; $\lambda'-\lambda$ is the range of wave-length in the mth order which lies between the mth and $(m+1)$th orders of λ.

$$m\lambda' = (m+1)\lambda,$$

$$\therefore \lambda' - \lambda = \frac{\lambda}{m}.$$

The minimum resolvable difference, $\delta\lambda$, is λ/mn, from which we see that a grating will resolve a difference of wave-length $1/n$th of the range of wave-length comprised within the width of an order.

In practice it is impossible to obtain a line source of light of infinitesimal width. A narrow slit at the focus of a collimating lens is usually employed. The spectral "lines" are in reality images of this slit and have therefore a finite width apart from the broadening due to diffraction. Such lines have to be further apart in order to be resolved. The ratio of the resolving power actually obtained with a given slit width to the theoretical resolving power mn is termed the *Purity* of the spectrum.

§ (7) Overlapping Spectra.—Since the number of elements is without effect on the dispersion in a given direction, the resolving power in a given direction, the width of the spectral lines, or, provided there are more than about a dozen elements, the relative intensities of the secondary maxima, it may not be obvious what advantage accrues from using a large number. There are several advantages. We saw earlier that in order to have light in oblique directions the width of each elementary slit must be small. If we only had a few of such slits the grating would be practically an opaque screen. For the sake of brightness, therefore, it is desirable to put as many slits into the available space as possible.

There is another important consideration, however, due to the overlapping of spectra of different orders. From the grating law the mth order spectrum of wave-length λ is formed in a direction such that

$$\Delta(\sin\theta - \sin i) = m\lambda.$$

In general, therefore, if $m_1\lambda_1 = m_2\lambda_2$, the wave-length λ_2 in the spectrum of m_2th order will coincide with λ_1 in the m_1th order. Thus the blue end of the second order spectrum ($\lambda = 0{\cdot}4\mu$) will overlap the red end ($\lambda = 0{\cdot}8\mu$) of the first order. Overlapping becomes progressively more serious in higher orders. If λ_1 and λ_2 are the extreme wave-lengths at either end of the spectrum to which the receiving apparatus (eye, photographic plate, or whatever it may be) is sensitive, a wave-length λ in the mth order will be overlapped by λ_1 in the m_1th order and λ_2 in the m_2th order, where $m_1\lambda_1 = m_2\lambda_2 = m\lambda$. All orders between m_1 and m_2 will clearly be represented at the same place by wave-lengths between λ_1 and λ_2, so that all orders from m_1 to m_2 inclusive, *i.e.* $m_2 - m_1 + 1$ orders, overlap at this point. Since

$$m_2 - m_1 + 1 = \frac{m\lambda}{\lambda_1\lambda_2}(\lambda_2 - \lambda_1) + 1, \qquad (8)$$

we see the importance of keeping m as small as possible if we wish to avoid confusion with polychromatic spectra. To reduce the order of the spectrum found in a given direction we must reduce the grating space Δ, that is, we must increase the number of elements in a given length of grating. Thus, although the resolving power does not depend on n, we obtain it in spectra of lower order, and thereby minimise the difficulties of overlapping, if we make n large.

§ (8) Distribution of Light in Spectra: Effects of Groove Form.—In what has gone before, we have assumed for convenience

that each element of the grating radiates uniformly in all directions. On this basis we found that the maximum intensities of the spectrum lines of all orders were equal, and that the secondary maxima had certain intensities relative to these. Since both primary and secondary maxima are extremely narrow, the eye does not appreciate their intrinsic brightness (light per unit area) but integrates this over the whole width of the lines and appreciates their total candle power;[1] consequently the relative apparent brightness of primary and secondary maxima is the ratio of the area of the intensity curves and not of their maximum ordinates. If we introduce the factor 2, the relative width of a primary and secondary band, and assume the primaries to be twice as bright, relative to the secondaries, as the ratio of the maximum ordinates obtained in preceding paragraphs, we shall obtain a fair approximation to the truth.

This phenomenon accounts for the apparent increase in the relative brightness of the secondary maxima when the aperture is reduced. In such circumstances the width of the maxima may increase until they occupy quite appreciable areas. The eye then measures their intrinsic brightness, and the secondaries seem twice as bright relatively to the primaries as when they are narrow lines. The effect of contrast also helps to subdue the secondaries when they are close to the primaries. What we have just said may explain the common conception that increasing the number of grating elements reduces the intensity of the secondary maxima. Such conclusions are usually drawn from experiments in which the aperture of a grating is varied. As we saw earlier, the relative maximum intensities are independent of the number of elements except when this is reduced to less than 20, and even then the difference is trifling. All the phenomena can be observed, as was pointed out in § (4), without any grating at all.

As regards the relative intensities of lines in different directions, these are clearly proportional to the ordinates of the intensity curve for the diffraction by the individual slits of the grating (see § (2)). In the case of the simple grating which we have so far considered, consisting of very narrow slits in an opaque screen, the central image would be of maximum brightness, while the spectra on either side would diminish symmetrically in brightness as the order increased. In practice, however, such a simple grating is not obtainable. Gratings are usually produced by ruling fine grooves on glass or metal with a diamond point. The technique of their manufacture is dealt with in another article,[2] from which the reader will find that the grooves are usually of V shape, and, for the best results, are cut sufficiently deep to leave none of the original surface remaining. The ruled surface is in fact a ploughed field in miniature. When the ruling is successful the sides of the groove are reasonably smooth, and act, in the case of metal gratings, as reflecting surfaces. Considering one side of these grooves the incident light would be reflected, in the absence of diffraction, in a certain direction in accordance with the ordinary law of reflection. As in the case of a simple slit, the diffraction spectra of the First Class are symmetrical about the direction which the rays would take in the absence of diffraction. If the ruling is narrow the central band occupies a wide angle, but its centre of symmetry and of brightness is in the direction of geometrical reflection. Similar considerations apply to the other sides of the grooves, the direction of maximum intensity being, of course, quite different for these. Each ruling therefore provides diffracted light strongly concentrated in two directions, and spectra lying near either of these directions will be relatively strong. This property is utilised to rule gratings which will concentrate the light in the spectrum of a particular order. The angle which a side of the groove must make with the plane of the grating in order that the central diffraction band from that side will lie in the same direction as the required spectrum is calculated, and the ruling diamond is set to give a groove of this form. An important experimental investigation on the effect of groove form on the intensity of different orders has been made by Trowbridge and Wood, while various writers have discussed the subject theoretically (References 10, 11, 12, 13, and others).

§ (9) Absent Spectra.—If the ruling is of appreciable width, the central diffraction band does not extend over the whole angular field, and first class minima will occur in certain directions. The light from each of the rulings is zero in those directions; consequently their joint action is zero, and any spectra which would otherwise be formed in such directions are absent. It is easy to deduce which spectra will be absent when assumptions are made as to the type of ruling and the method of using the grating, but such calculations lose in practical value from the difficulty of predicting or determining the precise form of a ruling. In general the orders of absent spectra are multiples of the first absent order. Thus, if the 2nd order is absent, the 4th, 6th, etc., are also absent; if the 3rd is absent, so also are the 6th, 9th, etc.; if the 4th is the first that is absent, then the 8th, 12th, etc., will be absent. It does not follow that the same orders are absent on both sides of the central image, since the centre of symmetry of the first class system does

[1] See article on "The Eye."
[2] "Diffraction Gratings, The Manufacture and Testing of."

not usually coincide with this in the case of a ruled grating. Not only so, but the spectra on opposite sides of the central image may be mainly due to light of the first class systems diffracted by the different sides of the groove, and are really due to different grating elements altogether, although the grating space is the same for both sets. With this brief notice the subject of absent spectra must be left. The reader desirous of further information must consult the references (2 and 3) in which the simple case of a grating consisting of transparent slits in an opaque screen, used at normal incidence, is treated at length.

II. Types of Grating

§ (10) Plane Grating.—This, the simplest type of optical grating, is usually produced by ruling parallel lines on optically flat surfaces by means of an accurate dividing engine.[1]

Two forms require mention, namely *transmission gratings*, in which the rulings are on glass or other transparent material, and *reflection gratings*, in which the rulings are on brightly polished metal. Transmission gratings are much more difficult to rule than reflecting gratings because of the hardness of the glass and the liability to change of form, if not complete breakdown, of the ruling point. For this reason reflecting gratings are almost invariably used for accurate work. On the other hand, small transmission gratings, suitable for the use of students, can be made very cheaply by taking a cast in celluloid of a good grating and mounting this on a glass plate. Such film replicas are sometimes of excellent quality. Photographic processes have also been used with some success. For accurate work, however, it is necessary to use a machine-ruled grating, as slight unevenness in shrinkage of the films on drying renders replicas of little use for research purposes.

A description of the various methods in which such gratings may be used is given in another article,[2] and need not be described here.

§ (11) Concave Grating.—Rowland discovered that if the grating were ruled on a concave spherical surface instead of a plane it produced focussed images without the use of a lens. This is not due to any new property acquired by the rulings but simply to the image-forming properties of a concave mirror superposed on the ordinary grating action.

The mirror causes a beam of light, diverging from the slit, to converge towards a real image in accordance with the ordinary laws of reflection. The effect of the grating is to deviate the rays of the convergent beam into a series of lateral spectra on either side of the geometrical image in accordance with the principles we have already discussed. It simplifies the conception of the concave grating if this is kept quite clearly in mind; the grating action is responsible for the presence of light in the directions given by the grating law, but the curvature of the wave front which produces the focussed images in these directions is due to the focal properties of the mirror.

Except for objects on the axis, a concave mirror produces astigmatic images. A point source so situated that the light is incident obliquely on the mirror gives rise to two focal lines respectively perpendicular and parallel to the plane of incidence. If instead of a point source we have a slit perpendicular to the plane of incidence, an image of the slit, drawn out in the direction of its length, *but not broadened*, will be formed at the first of these focal lines. It is this astigmatic image with which we are concerned, since it is the nearest approach to an "image" of a slit which the mirror will give. If S, *Fig.* 4, is the slit

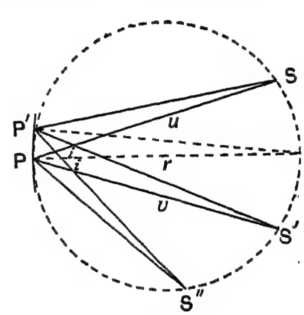


Fig. 4.

situated at a distance u from a concave mirror of radius r, the image is formed in a direction equally inclined to the normal and at a distance v in accordance with the ordinary law for oblique reflection by such mirrors, viz.

$$v=\frac{ur\cos i}{2u-r\cos i}.$$

Wherever the object may be situated, an image will be formed in the position given by this law. The quality of the image is not, however, equally good for all positions of S. Like all optical instruments a mirror suffers from spherical aberration. In the case of a spherical surface this aberration is least, being practically absent, if the object and image are equally distant from the mirror. In this case $u=v=r\cos i$, that is, the object (and also the image) is situated on the circumference of a circle with PC, the radius of the mirror, as diameter. We may term this the *focal circle*;

[1] See article on "Diffraction Gratings, The Manufacture and Testing of."
[2] "Wave-lengths, Measurement of."

the slit must be situated on this circle if the image is to be free from spherical aberration.

Let us now consider the effect of the grating action on the rays converging to S′. Part of the light will be thrown into a series of lateral spectra at inclinations, θ, to the normal CP such that $\Delta(\sin\theta - \sin i) = m\lambda$ in accordance with the usual grating law. Since S, C, P, and any point P′ on the grating are concyclic,[1] $S\hat{P}'C = S\hat{P}C$, that is to say, the angle of incidence is constant over the grating. All rays in the beam are therefore deviated to the same extent, so that the convergence in the lateral spectra is the same as in the central beam. Thus if S″ is one of the spectra,

$$P\hat{S}''P' = P\hat{S}'P',$$

i.e.
$$\frac{PP'\cos\theta}{PS''} = \frac{PP'\cos i}{PS'},$$

or
$$PS'' = v\frac{\cos\theta}{\cos i} = r\cos\theta.$$

Hence if the slit is situated on the focal circle not only the central image but also the spectral images are situated on this circle.

In the general case in which S is not situated on the focal circle, the position of the spectral images is still obtained to a close approximation by regarding the central beam as deviated without change of convergence, so that if v is the distance of the central image,

$$PS'' = v\frac{\cos\theta}{\cos i},$$

$$PS' = \frac{ur\cos\theta}{2u - r\cos i} \quad . \quad . \quad . \quad (9)$$

This, however, is not quite accurate, because in the general case the angle of incidence on the grating, and consequently the deviation of the rays by the grating action, varies across the surface. The convergence of the spectral beams is therefore not exactly equal to that of the central beam, although each ray is simply deviated in accordance with the relation $(\sin\theta - \sin i) =$ constant, and the points of convergence such as S″ are nearer or further from the grating than indicated by (9). The correct formula[2] is more readily obtained on somewhat different lines than those followed above, which have been adopted in order to show that the focal effect is purely the ordinary property of the concave mirror, and that the effect of the grating, whether plane or concave, is merely to deviate the light into lateral spectra in accordance with the grating law.

The error of formula (9) is very slight in any case of practical importance. Since the grating merely deviates the rays, the aberration which characterises the central image if the slit is far from the focal circle also characterises the spectral images, so it is necessary for good definition to keep to the concyclic arrangement.

True focal effects of gratings, which do not depend on mirror action, may be encountered if the grating space is not constant but varies according to certain laws. These effects are discussed in the article on the "Manufacture and Testing of Diffraction Gratings" and need not be referred to further here. In any case they are freak effects and cannot be produced at will on any type of ruling engine at present in use.

Discussions of the aberrations of concave gratings will be found in references 26, 27, 28 and others.

In Rowland's method of mounting the grating the slit is placed at such a point on the circumference of the circle that the spectrum is formed at C, *i.e.* $\theta = 0$. This arrangement gives a normal spectrum, *i.e.* a spectrum in which equal distances on the photographic plate (which is bent to be tangential to the focal circle) represent equal increments of wavelength. We saw earlier (§ 7) that the dispersion of a grating, $d\theta/d\lambda = m/\Delta\cos\theta$. The rate of variation of dispersion from one part of the spectrum to another is

$$\frac{d}{d\theta}\left(\frac{d\theta}{d\lambda}\right) = \frac{-m\sin\theta}{\Delta\cos^2\theta},$$

and is zero if $\theta = 0$.

In the apparatus used by Rowland the grating and photographic plate are mounted opposite each other at the extremities of a rigid beam GP, *Fig.* 5, P being at the centre of curvature of the grating. The two ends of the beam can be moved along rails, *S* A and *S* B at right angles to each other. The slit is mounted at *S*. It is obvious that whatever position GP occupies *S* is on the semicircle with GP as diameter. The instrument when once adjusted is therefore always in focus. All that is required is to move the bar GP until the required spectrum occupies the plate. Full details of the use of Rowland's mounting are given in Kayser's *Handbuch der Spectro-*

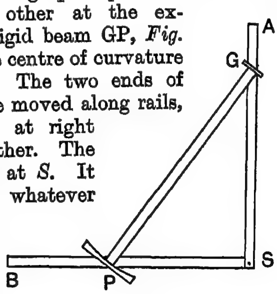


FIG. 5

[1] The diameter of the grating is never such a large fraction of its radius of curvature that the distance of any point on it from the focal circle is appreciable.

[2] This formula is

$$PS'' = \frac{ur\cos^2\theta}{ur(\cos i + \cos\theta) - r\cos^2 i}.$$

The condition for the formation of an optical image is that terms involving the square of the aperture should vanish in the expression for the retardation of rays from different parts of the aperture. The above formula follows at once in the case of a mirror if θ is different from i. When $\theta = i$, that is for the ordinary catoptric image, it reduces to the usual formula connecting u and v.

scopie, vol. i. A defect of Rowland's mounting is that it takes up very considerable space. An arrangement in which the plane of the apparatus is vertical, so that all except the slit and camera are contained in a pit beneath the floor, has been described by A. S. King (20). Eagle (21) has described a most useful mounting for a concave grating in which the centre of the plate coincides (virtually) with the slit. Full details and a critical comparison of its advantages and disadvantages as compared with Rowland's mounting are contained in the paper. Its two main advantages are that higher orders can be used and that the astigmatism of the spectrum is very much less. The outstanding defect of concave gratings is the astigmatism of the images, particularly in the higher orders, for although this does not affect the sharpness of the definition the brightness is greatly impaired, which is a serious matter in photographing faint spectra. With Eagle's mounting much shorter exposures are required than with Rowland's arrangement.

§ (12) THE ECHELETTE.—For work in the infra-red region R. W. Wood (10) constructed gratings of wide spacing, $\Delta = \cdot 0123$ mm. These were ruled on a gilt copper plate, using the 120° angle of a carborundum crystal. In the majority of cases this was mounted to cut grooves one face of which made an angle of about 20° with the original surface. With normal incidence these facets give a concentration of light with the centre at 40° from the normal. With visible light such gratings throw the energy into a small group of spectra near the 15th order on one side of the normal. They show no central image nor any spectra to the other side. According to Wood,[1] "The gratings behave, with infra-red radiation of wave-length above, say, 3μ, as almost ideally perfect gratings; that is, they give spectra similar to what we should have with an ordinary grating which threw practically all of the light into one or two orders on one side of the central image." For this wave-length, 3μ, the brightest spectrum will be of the 2nd or 3rd order. The useful spectra for wave-lengths in this neighbourhood are not therefore of high order; so that the resolving power is much less than for visible light. This is no actual loss, however, because the relatively large width of the radiometric receiver makes it impossible to utilise such high resolving powers as in the visible and photographable regions of the spectrum.

Wood has termed these gratings "*Echelettes*" because, in their relatively large phase-step, about 15λ for visible light, they lie between the ordinary grating and the Echelon, which we are now about to describe.

§ (13) THE ECHELON.—This instrument, which was devised by Professor Michelson (29), is as unlike the traditional grating as possible. It consists of a number of glass plates, of precisely the same thickness and plane parallel to a high degree, arranged like a flight of steps, or *in echelon*, whence the name. The width of the steps is made as accurately equal as possible. If the light is incident normally on the broadest plate, *Fig.* 6, some of it emerges from each of the steps which act as rectangular apertures and diffract the light over a very small angle. The light emerging from any step has traversed one plate more or less than from the next, so there is a constant phase-step depending on the thickness and refractive index of the plates. Let $a\,e$ and $c\,e'$ be rays diffracted in a direction θ from the outer edges of two adjacent steps. Let $c\,f$ and $b\,g$ be perpendicular to $a\,e$. The difference in optical path is obviously the difference between $b\,c$ in glass and $a\,f$ in air. Let t be the thickness and β the width of a step, then

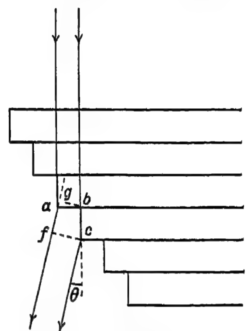


FIG. 6.

$$af = gf - ga = t\cos\theta - \beta\sin\theta = t - \beta\theta,$$

since θ is necessarily very small.

Thus the phase-step is

$$\mu t - (t - \beta\theta) = (\mu - 1)t + \beta\theta.$$

A spectrum line of the mth order will therefore be formed when

$$(\mu - 1)t + \beta\theta = m\lambda. \qquad (10)$$

Since θ is very small m is approximately $(\mu-1)t/\lambda$. For a plate of thickness 2 cm. $(\mu - 1)\,t$ is about 20,000 wave-lengths for green light, and the order of the spectrum is 20,000 or so.

The dispersion,

$$\frac{d\theta}{d\lambda} = \left(m - t\frac{d\mu}{d\lambda}\right)\Big/\beta,$$

$$= \left\{\frac{(\mu-1)t}{\lambda} - t\frac{d\mu}{d\lambda}\right\}\Big/\beta,$$

$$= \frac{bt}{\lambda\beta},$$

where $$b = \left\{(\mu - 1) - \lambda\frac{d\mu}{d\lambda}\right\}$$

and is purely a property of the glass. The total aperture is $n\beta$ where n is the number of plates. The minimum angle resolvable is therefore $\lambda/n\beta$.

[1] *Physical Optics*, 1911, p. 229.

The minimum wave-length difference resolvable is

$$\delta\lambda = \frac{\lambda}{n\beta} \cdot \frac{d\lambda}{d\theta} = \frac{\lambda^2}{nbt}.$$

Hence the resolving power, R,

$$= \frac{\lambda}{\delta\lambda} = \frac{nbt}{\lambda}. \quad . \quad . \quad . \quad . \quad (11)$$

But $\qquad m = (\mu - 1)\frac{t}{\lambda},$

$$\therefore R = mn\frac{\{\mu - 1 - \lambda(d\mu/d\lambda)\}}{(\mu - 1)}.$$

Thus the effect of the dispersion of the glass is to increase the resolving power (since $d\mu/d\lambda$ is $-ve$) from that given by the ordinary grating formula $R = mn$. In the case of a particular echelon described by Baly,[1] $\mu - 1$, in the neighbourhood of the sodium line, was 0·5746, while $\mu - 1 - \lambda(d\mu/d\lambda)$ was 0·6165.

For an echelon of 20 plates 2 cm. thick, the resolving power is a little over 400,000 for visible light. Such a grating would therefore resolve lines differing in wave-length by only one four-hundredth of the difference of the sodium lines.

This large resolving power, obtained with a relatively small aperture,[2] may appear to contradict the conclusion of §§ (6) and (7), that the resolving power for a given aperture was independent of the grating space. We were there concerned with ordinary gratings in which the phase retardation between adjacent elements depends simply on the obliquity of the light.

We must not confuse the width of the echelon step with the grating space of an ordinary grating. The width of the step corresponds to the width of one of the elementary apertures of the grating and determines the distribution of light in the field in accordance with § (2), but the analogue in the case of the echelon to the grating space of the ordinary grating is the separation of the apertures *parallel to the length of the echelon*, which introduces the enormous phase-step $(\mu - 1)t$ (approx.) between rays from corresponding points of the different apertures. This results in a dispersion very much greater than anything obtainable with an ordinary grating, so that for a given total aperture, and therefore given resolving power *in angle*, the spectral resolution is greatly increased. An ordinary grating of 20,000 lines per inch would require to be ten inches long to give the same spectral resolution in the second order as we obtain with an echelon of under an inch.

The difficulties of ruling such a grating are very considerable, and very few of this length have been produced at the present time, whereas the echelon we have mentioned is by no means an extreme case.

The angular field within which all the phenomena are compressed is the angular width of the central diffraction band from a single aperture. This is $2\sin^{-1}(\lambda/\beta) = 2\lambda/\beta$.

[1] "Spectroscopy," chapter vi.

[2] β is usually about 1 mm., so $n\beta$ is 2 cm. in the case we are discussing.

The angular separation of successive orders will be given by $d\theta/dm$ since m is large.

From (10) $\qquad \dfrac{d\theta}{dm} = \dfrac{\lambda}{\beta}.$

Thus the distance between two orders is half the width of the total region in which light is available. There will thus in general be two orders visible, of different intensities unless they happen to be symmetrical with respect to the centre of the diffraction maximum. By slightly altering the angle of incidence one of the orders may be brought to the centre, when it will be of maximum brightness; the spectra of adjacent order on either side then coincide with the diffraction minima, and are invisible or faint. These two positions are termed positions of single or double order.

Though not more than two orders of *one wave-length* are in the field at once, if another wave-length is present this will appear in some other order. The mth order of wave-length λ will coincide with the $(m \pm p)$th order of wave-length $\lambda m/(m \pm p)$, where p is any integer 1, 2, 3, . . ., etc. Since m is large, a very slight range of wave-length indeed will give rise to so many overlapping spectra of different orders that complete confusion will occur. With the sodium lines, for instance, if D_1 is in the 20,000th order D_2 may be in the 20,020th. The lines may overlap or lie between one another, or, in fact, have any relative position whatever, depending on the exact thickness of the plate and the exact angle of incidence.

Thus very homogeneous radiation must be employed or the appearances presented are unintelligible. In practice, therefore, the light is always analysed by means of an auxiliary spectroscope, usually a prismatic instrument, though sometimes a grating is employed, and only that radiation is allowed to enter the echelon which it is desired to examine.

Another drawback to the instrument is this. The closest doublet which a grating will resolve is, as we saw in § (6), separated by $1/n$th of the separation of adjacent orders. Since the useful field is the space between two orders we can only examine a range of wave-length of about n times the resolvable minimum. Since, in the case of the echelon, n is so small, this limitation is serious. It is clearly advantageous for a given resolving power to use a large number of thin plates rather than a smaller number of thick ones. However, owing to practical difficulties of construction and to loss of light by reflection at interfaces, it is not feasible to use a very large number of plates. There is little gain in going beyond 30, while most echelons have under 20.[3]

[3] Messrs. Adam Hilger, as the result of improved methods of construction, have recently put echelons of fifty-six plates on the market.

Since a broadened line, or the components of a multiple line may extend over the whole width of the first class maximum, the apparent brightness will be misleading. Correction must be made for the distribution of intensity within the field. The formula of § (2) reduces in this case to

$$I = I_0 \frac{\sin^2 (\pi\beta/\lambda)\theta}{\{\pi(\beta/\lambda)\theta\}^2},$$

where I_0 is the intensity at the centre of the diffraction band. An experimental investigation of the agreement of practical results with this formula has been made by Burger and van Cittert (Ref. 33), who also pointed out that owing to that side of a spectral maximum which is further from the centre of the field being more weakened than the other, the position of the maximum is appreciably displaced towards the centre of the field. They found errors amounting in some cases to 0·5 per cent of the width of an order. This displacement appears to have been generally overlooked by users of the instrument. It is analogous to the displacement of the secondary maxima on account of their rapidly diminishing intensity, to which attention was called in § (2).

When all its difficulties have been taken into account, however, the Echelon remains a beautiful and powerful weapon of physical research. It is easy to use, requiring little adjustment, and as the pioneer among spectroscopes of extreme resolving power, has been responsible for the opening up of important fields of investigation previously inaccessible. The actual use of the instrument is dealt with in another article.[1]

§ (14) Defective Gratings.—We have hitherto assumed ideal gratings and investigated their properties. In practice gratings may suffer from various defects. Principal among these are departures from exact equality of the grating space over the whole grating. The effects of such errors are dealt with in another article,[2] and we need only mention here that any periodic variation in the ruling virtually constitutes a second grating of which the grating space is equal to the periodicity. Thus a spectrum line is flanked with a series of faint companions which are simply lateral spectra produced by the second grating. Such companion lines are termed *ghosts*. Variations in optical thickness of the plates of an echelon may also give rise to ghosts. Aberrations of another character may also be found in echelons due to the clamping of the plates (30). Stansfield and Walmsley (32) described a case in which an asymmetrical distribution of light in the secondary maxima resulted from this cause.

A useful treatment of imperfect gratings has been given by Sparrow (35).

§ (15) Conclusion.—In an article of this length it has been quite impossible to deal exhaustively with the subject of diffraction gratings, either in the theoretical or practical aspects. It has not been possible to deal with the practical methods of mounting employed, and for such information the reader is referred to Kayser's *Handbuch der Spectroscopie*, vol. i., to Baly's *Spectroscopy*, and to several of the papers mentioned in the bibliography. On the theoretical side, space has forbidden reference to the state of polarisation of the diffracted light from the grating elements (37 to 43). To repair these omissions and supplement the information contained in the article, a bibliography is appended which, though far from complete, will be found useful by those desirous of pursuing the subject further.

J. G.

[1] "Modern Spectroscopy."

[2] "Diffraction Gratings, The Manufacture and Testing of."

Bibliography

1. Rowland, *Phil. Mag.* (5), 1893, xxxv. 397; *Astronomy and Astrophysics*, 1893, xii. 129.
2. Rayleigh, *Ency. Brit.*, articles on "Wave Theory" and "Optics."
3. Rayleigh, *Phil. Mag.*, 1874, xlvii. 81, 193, "On the Manufacture and Theory of Diffraction Gratings."
4. Rayleigh, *Proc. Roy. Soc.* A, 1907, lxxix. 399, "On the Dynamical Theory of Gratings."
5. Kimball, *Phil. Mag.* (6), 1903, 30, "Note on the Application of Cornu's Spiral to the Diffraction Grating."
6. Wood, *Phil. Mag.*, 1907, xiv. 477, "A Simple Treatment of the Secondary Maxima of Grating Spectra."
7. Rayleigh, *Collected Researches*, i. 413 *et seq.*, various papers on the Resolving Power of Optical Instruments.
8. Porter, *Phil. Mag.*, 1908, xv. 762, "On the Effect of Position of the Grating (or Prism) on the Resolving Power of a Spectrum."
9. Wadsworth, *Phil. Mag.* (5), 1897, xliii. 317, "On the Resolving Power of Telescopes and Spectroscopes for Lines of Finite Width."
10. Wood, *Phil. Mag.*, 1910, xx. 770, "Echelette Gratings for the Infra Red"; Trowbridge and Wood, *ibid.* p. 886, "Groove Form and Energy Distribution of Diffraction Gratings"; *ibid.* p. 898, "Infra-red Investigations with the Echelette Grating."
11. Trowbridge and Crandall, *Phil. Mag.*, 1911, xxii. 534, "Groove Form and Energy Distribution of Diffraction Gratings."
12. Anderson and Sparrow, *Ast. Journ.*, 1911, xxxiii. 338, "On the Effect of Groove Form on the Distribution of Light by a Grating."
13. Trowbridge, *Ast. Journ.*, 1911, xxxiv. 317, "Concerning the Echelette Grating."
14. Kayser, *Handbuch der Spectroscopie*, i.
15. Rayleigh, *Phil. Mag.*, 1881, xi. 196, "On Copying Diffraction Gratings and some Phenomena connected therewith."
16. Rayleigh, *Nature*, 1896, liv. 332, "The Reproduction of Diffraction Gratings."
17. Anderson, *Ast. Journ.*, 1910, xxxi. 171, "Glass and Metallic Grating Replicas."
18. Wood, *Phil. Mag.*, 1906, xii. 585, "The Intensification of Glass Diffraction Gratings and the Diffraction Process of Colour Photography."
19. Ebert, *Wied. Ann.*, 1889, xxxviii. 489, "Zwei Formen von Spectrographen."
20. King, *Ast. Journ.*, 1914, xl. 205, "A Vertical Adaptation of the Rowland Mounting for a Concave Grating."
21. Eagle, *Ast. Journ.*, 1910, xxxi. 120, "A New Mounting for a Concave Grating."
22. Wadsworth, *Ast. Journ.*, 1895, ii. 370; *ibid.*, 1896, iii. 46.
23. Ames, *Phil. Mag.* (5), 1889, xxvii. 369.
24. Adeney and Carson, *Proc. Roy. Soc. Dublin* (1), 1898, viii. 711, "On the Mounting of the Large Rowland Spectrometer in the Royal University of Ireland."
25. Birge, *Ast. Journ.*, 1916, xliii, "Notes on the Use of Concave Gratings."

26. Wadsworth, *Phil. Mag.*, 1903, vi. 119, "On the Aberrations of the Concave Grating used as an Objective Spectroscope."
27. Rowland, *Phil. Mag.*, 1883 (5), xvi. 197.
28. Glazebrook, *Phil. Mag.*, 1883 (5), xv. 414; *ibid.* xvi. 337.
29. Michelson, *Ast. Journ.*, 1898, viii. 37; *American Journ.* (4), 1898, v. 215; *Trans. Camb. Phil. Soc.*, 1899, xviii. 316.
30. Twyman, *Proceedings Optical Convention*, 1905, p. 50, "The Michelson Echelon Diffraction Grating and the Lummer Parallel Plate."
31. Stansfield, *Phil. Mag.*, 1909, xviii. 371, "The Echelon Spectroscope and its Secondary Action, and the Structure of the Green Line of Mercury."
32. Stansfield and Walmsley, *Phil. Mag.*, 1912, xxiii. 25, "Changes in Diffraction Spectra Produced by Aberration, and the Aberration of the Echelon Grating."
33. Burger and van Cittert, *Proc. Amsterdam Akad.*, 1920, xxiii. No. 5, "Measurements on the Intensity of Spectral Lines by the Aid of the Echelon."
34. Michelson, *Ast. Journ.*, 1903, xviii. 278, "On the Spectra of Imperfect Gratings."
35. Sparrow, *Ast. Journ.*, 1919, xlix. 65, "Theory of Imperfect Gratings."
36. Allen, *Phil. Mag.*, 1903, vi. 559, "The Effect of Errors in Ruling on the Appearance of a Diffraction Grating."
37. Fizeau, *Ann. de Chimie*, 1861, lxiii. 385.
38. Rayleigh, *Phil. Mag.*, 1907, xiv. 350-359, "On the Light Dispersed from Fine Lines Ruled on Reflecting Surfaces or Transmitted by very Narrow Slits."
39. Du Bois and Rubens, *Phil. Mag.*, 1911, xxii. 322, "On Polarisation of Heat Rays by Wire Gratings."
40. Voigt, *Gött. Nachr.* Math.-Phys. Kl., 1911, p. 40.
41. Frolich, *Wied. Ann.*, 1881, xiii. 133, "Neue Modification des Lichtes durch Reflexion an engen Metallgittern."
42. Frolich, *Wied. Ann.* 1882, xv. 577, "Experimentaluntersuchungen über die Intensität des gebeugten Lichtes."
43. Ingersoll, *Ast. Journ.*, 1920, li. 129, "Polarisation of Radiation by Gratings."

Diffuser, a surface which scatters the light falling on it so that the light reflected or transmitted is distributed according to the cosine law. See "Photometry and Illumination," § (51).

Diffusion-area Instrument: a type of ophthalmic instrument which depends upon diffusion areas for the extinction of refraction. See "Ophthalmic Optical Apparatus," § (10).

Diffusivity, Visual. See "Eye," § (19).

Dip of Sea Horizon: method of measurement. See "Navigation and Navigational Instruments," § (24).

Dipping Refractometer: a direct-reading refractometer of limited range used for specific purposes. See "Spectroscopes and Refractometers," § (16).

Direct Lighting. See "Photometry and Illumination," § (71).

Direct Vision Spectroscope. See "Spectroscopes and Refractometers," § (19).

Direction, Perception of Sound. See "Sound," § (57) (v.).

Directional Wireless Telegraphy, its application to position fixing at sea. See "Navigation and Navigational Instruments," §§ (25), (26).

Disc, Rayleigh, used as a sound detector. See "Sound," § (56).

Discharge Tube, the phenomena of the vacuum. See "Radiology," § (3).

Disintegration Theory of radioactivity, put forward in 1903 by Rutherford and Soddy to explain the continuous production of radioactive matter. See "Radioactivity," § (7).

Dispersion, Irrationality of. See "Optical Glass," § (2).

Dispersion, Rotatory. See "Polarised Light and its Applications," § (21) (ii.).
Biot's Investigations of. See "Quartz, Optical Rotatory Power of," § (3) (i.).
Drude's Theory of. See *ibid.*, § (3) (ii.).

Dispersion, Rotatory, Experimental Investigation, Drude's Theory of. See "Quartz, Optical Rotatory Power of," § (4).

Dispersion in grating spectra. See "Diffraction Gratings, Theory of," § (6).

Dispersion of Light. The separation of a beam of light into its constituent colours, each due to a wave of definite length. See "Spectroscopes and Refractometers"; also "Modern Spectroscopy."

Distortion of Optical Images. See "Telescope," § (3). See also "Lenses, Aberrations of," § (5); "Optical Calculations," § (18).

Diurnal Variation of Daylight. See "Photometry and Illumination," § (74).

Diversity Factor (in Illumination): the ratio of the maximum to the minimum illumination found within a given area. See "Photometry and Illumination," § (70).

DIVIDED CIRCLES

I. Historical

§ (1) General.—The graduated or divided circle is the basis of all instruments for measuring angles. The earlier circles were divided by hand, and in order to get any degree of accuracy they had to be of large diameter. During the last 150 years there has been a great development in the art, and it is now possible to obtain circles of a foot or 18 in. diameter in which no division has a greater error than 2 seconds of arc, while the probable error of any division is less than half a second. The first notable example of a divided circle was a mural circle of 8 ft. radius divided by George Graham for Greenwich Observatory in 1725. Two circles of 96·85 and 95·8 in. were marked off by beam compasses, the inner, or working circle being graduated to degrees and twelfths of a

degree. The chord having the same length as the radius gave the 60° point, with the points 0° and 60° as centres, and a distance on the beam compass very nearly equal to the chord of 30°; two light dots were made on the arc, and the distance between these was bisected by hand, using a magnifying glass, thus giving the 30° mark. The chord of 30° set off from the 60° point gave the 90° point. Each 30° arc was then bisected in a similar manner, and the resulting divisions trisected, thus dividing the quadrant into 5° spaces. These were then divided into 5 parts, giving single degrees, and the twelfths of a degree were found by further bisection and trisection. The outer circle was divided into 3 parts as before, and each of these into 512 parts by continued bisection. This method was considered more accurate than the first, and the resulting divisions were used as a check on the inner circle.

John Bird[1] in 1767 divided a similar circle. He first prepared a lineal scale of equal parts, and having obtained the 60° point by laying off the chord equal to the radius, 30° and 90° were obtained by laying off the chord of 30° from the 60° point, the length of this chord being obtained by calculation and measurement from his scale. From 60° the chord of 15° was laid off giving 75°, and from this the chord of 10° 20′ gave 85° 20′, which was checked by the chord of 4° 40′ from 90°; $85° 20' = 5' \times 2^{10}$, hence the further subdivision of this can be found by continued bisection. The remainder of the quadrant beyond 85° 20′ contains 56 divisions of 5′ each; the chord corresponding to 64 such divisions was laid off, and subdivided by continued bisection as before.

§ (2) Troughton.—Edward Troughton adopted another method for dividing (described in *Phil. Trans.*, 1809, as applied to a 4 ft. circle). The circle to be divided was first accurately turned on its inner and outer edges as well as on its face. A cylindrical roller was then made of such diameter that it revolved 16 times on its axis when rolled once round the outer edge of the circle. The roller was then divided as accurately as possible into 16 equal parts by lines parallel to its axis, and was mounted on a framework which could slide round the circle, the roller revolving by means of frictional resistance on the outer edge of the circle. Two microscopes attached to the frame served to observe the circle and the divisions on the roller respectively. By means of these microscopes the points of contact of the marks on the roller with the circle were observed and marked on the circle by dots; the circle was thus divided into 256 nearly equal parts. Two microscopes A and B were used to obtain the errors of these dots; they were placed so as to bisect dots 0 and 128 respectively; the circle was then turned so that A bisected 128. B should then bisect dot 0; if it did not, then half the movement of the cross-hairs necessary to bisect the dot gave the error of the dot, and this was measured on the micrometer screw. Microscope A was then again set over dot 0 and B over dot 64; the circle was then turned till A bisected dot 64. B should then bisect dot 128 ± the error already found for dot 128; half the residual error gave the error of dot 64. Similarly the error of dot 192 was found. By continued bisection in the same manner the errors of all the other dots were found. The final graduations were required at 5′ intervals, and to obtain these a sector with about four times the radius of the roller was mounted concentrically on the roller, so that it normally turned with the roller, but could be adjusted independently. The sector was divided as accurately as possible into divisions such that it would revolve one division on its axis for each 5′ which the frame was moved round the circle, $16\frac{7}{8}$ of such divisions corresponding to $\frac{1}{16}$ of the circumference of the sector, *i.e.* to the amount the sector would revolve as the roller moved between two consecutive dots on the circle. Sixteen such divisions were marked on the sector, together with an extra division at each end which was divided into $\frac{1}{8}$ths. Two microscopes H and K were carried by the frame, H reading on the circle and K on the sector; the frame was then adjusted so that H read dot 0, and the sector adjusted with its zero under K. The zero mark of the circle was then cut under H by a dividing knife, and the frame moved till the sector had revolved one division under K; the next line was then cut under H, and so on, until 16 divisions had been cut. Before cutting the 17th H was adjusted over dot 1, due allowance being made for its error as already determined, and the sector was set to division $-\frac{1}{8}$ under K; the graduations between dots 1 and 2 were then cut on the circle. In this way the errors of the dividing were controlled by the dots and prevented from accumulating.

§ (3) Copying.—When a circle has once been accurately divided it can be copied by placing the master circle and the blank concentrically on a table, and laying a straight-edge to each division of the master circle in turn, and cutting the division on the blank with a dividing knife.

§ (4) Dividing Engines.—Modern circles are invariably machine divided. The first notable dividing engine was made by Jesse Ramsden,[2] and an account of this machine was published by the Commissioners for Longitude

[1] John Bird, *The Method of Dividing Astronomical Instruments*. London, 1768.

[2] J. Ramsden, *Description of an Engine for Dividing Mathematical Instruments*. London, 1777.

in 1777. This and all later engines are made on the same general lines; the work to be divided is mounted concentrically on a circular plate which revolves on a vertical axis. This plate has teeth cut on its edge and is rotated by a worm; the worm is moved by a ratchet driven by a band wound round a cylinder, stops being provided to ensure the worm being turned the correct amount. In the earlier machines the band was pulled down by a treadle, the other end of the band being attached to a weight. On pressing down the treadle the plate was rotated by a definite amount and the divisions marked on the circle by hand. On releasing the treadle the ratchet allowed the weight to descend without rotating the worm. In later machines the cuts are made automatically, thus enabling the engine to be power driven.

In 1793 Edward Troughton completed a somewhat similar machine. Others have been made by Andrew Ross,[1] William Simms,[2] and others in this country and abroad.

II. Modern Dividing Engines

§ (5) Description.—One of the best and most modern of these was completed by George Watts in 1905. This machine and the methods used in constructing it will be described below.

It will be seen from what has been said that the principal aim is to construct an accurate worm and worm-wheel so that one revolution of the worm shall revolve the worm-wheel through a predetermined angle, no more, no less, with the smallest possible amount of deviation, and this must be true in any relative position of the worm and wheel. The only certain method of producing such a worm-wheel is to cut the teeth one by one from an accurately divided circle, and unless such a circle is available the first step is to divide one. In order to carry this out the following precautions were taken. The work of dividing and testing was carried out in a specially constructed room where the temperature could be maintained constant within one degree Fahr. The circle A (*Fig.* 1) to be graduated revolved smoothly upon an accurately machined axis B without the suspicion of any journal shake. The bearings of this axis were carried by a very rigid frame C, which was so designed as to furnish also a rigid circular table D, somewhat lower in height and a little larger in diameter than the circle to be graduated. This circular table was for the purpose of mounting micrometer microscopes at different positions round the circle. These microscopes should be of sufficient power to detect clearly an amount equal, in angular value, to the tenth part of a second on the circle. Some idea of the power necessary can be obtained from the fact that in

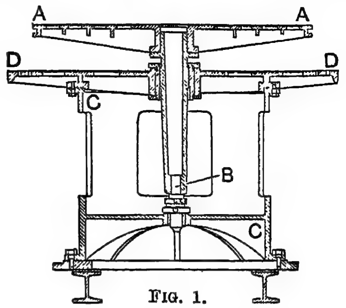


Fig. 1.

linear measure this amounts at 4 ft. diameter to about $\frac{1}{80000}$ inch. The design for the mounts of these microscopes must be very carefully considered from the point of view of temperature changes, as a cold draught striking one side of such a mount may produce errors which are by no means negligible.

Another essential piece of apparatus is a mechanically mounted steel or diamond cutter for cutting the actual divisions to be read by the micrometers. This cutter must be so mounted and swung that it can be lowered into contact with the circle, moved across it, and again raised mechanically without any direct impetus or control of the hand. Obviously also the moving parts of this mechanism must be free from the least shake, while at the same time perfectly free to move or swing with the least possible friction.

§ (6) Dividing the Master Circle.—The cutter mechanism was securely bolted down to the table D, and the first line ruled (0°). Then by actual measurements the position of the 180° line was determined, and this position brought as nearly as possible under the cutter. A micrometer microscope A was then mounted and adjusted exactly over the 0° line, *i.e.* directly opposite the cutter, and 180° line ruled; the circle was then revolved as nearly as possible through 90°, and microscopes B and C mounted over the marks already cut; they will thus be about 90° from A. Microscope B was then adjusted exactly over the zero mark and C over the 180° mark, and the circle turned till the zero mark was exactly under C; if the 180° mark had now appeared exactly under B, both marks and microscopes would have been exactly 180° apart, but if not, the error was shared equally by the positions of microscope B and line 180°. The total amount of this error was read on the micrometer, and the zero line again placed under the A microscope, but with its error allowed for on the A micrometer, and a fresh 180° line was ruled. This was again tested,

[1] *Trans. Soc. Arts*, 1830-31.
[2] *Memoirs of the Astronomical Society*, 1843.

and the process repeated until the two lines 0° and 180° had no appreciable error; the micrometer reading at A corresponding to the point exactly opposite the cutter was then also known. The zero line was then adjusted under B and the 90° line cut; 0° was then placed under C, when 90° should appear under A. As before, the error was read off, half this error being in the position of the 90° line, and half in the positions of B and C. A correct 90° line could therefore be marked, and the 270° line followed. By similar methods, and using more microscopes, accurate lines were cut first every 45° and then every 22½°. This gave 16 fundamental lines. By the same methods of trial and error microscopes D and E were placed 120° from A and the 120° and 240° lines cut. By placing each of the original 16 lines under each of the three microscopes A, D, and E in turn, three times 16, or 48, lines were obtained. On the same principle five microscopes were placed at equal distances round the circle, and more lines cut, thus giving 240 lines in all, or one every 1½°. So far the work was straightforward, if tedious; but these divisions had now to be divided into 18ths in order to obtain 5′ spaces.

For this purpose micrometers were mounted at 0°, 70° 30′, and 141°, and the two spaces included were trisected, *i.e.* intermediate micrometers were mounted at 23° 30′, 47°, 94°, and 117° 30′. By this means 720 divisions were obtained, or one every ½°. Three micrometers were then mounted at 0°, 60° 30′, and 121°, and the two spaces subdivided into three by micrometers mounted at 20° 10′, 40° 20′, 80° 40′, and 100° 50′. By this means 2160 divisions were obtained, or one every 10 minutes. Finally four micrometers were mounted at 0°, 40° 10′, 80° 20′, and 120° 30′, and the three spaces bisected by micrometers at 20° 5′, 60° 15′, and 100° 25′. This gave the required 4320 divisions, or one every 5 minutes.

§ (7) Testing. — The circle being thus divided the same geometric process was employed to ascertain the individual errors of the graduations, which were tabulated. An entirely fresh set of graduations was then cut, using a slightly different diameter of circle, the errors being eliminated, as far as possible, by making the appropriate allowances on the micrometers. The errors of this new circle were again determined and tabulated, and yet another circle divided. It was found that the individual errors were considerably diminished with each cycle of operations, and after a considerable number of such cycles, involving about six months of very tedious work, a ring of divisions was obtained the errors of which were known, and with one or two exceptions were well within ½ second of arc.

§ (8) Teeth Cutting.—The next step was to cut the teeth of the worm-wheel. In the place where the driving worm was finally to be mounted was placed a spindle carrying a multi-toothed V cutter of about 35° angle. The axis of this cutter was mounted in an inclined position, not parallel to the plane of the revolving table. The amount of this inclination was the mean angle of that portion of the thread of the worm that was to engage and drive the revolving table. Mechanical means were provided for revolving this V cutter and also for feeding it into the edge of the revolving table as required. There were also two or three warning devices for the purpose of indicating when the cutter was well out of contact with the revolving table so that the latter could be moved with impunity.

Seven micrometers were then mounted round the revolving table, and also, in a convenient position among them, a firm clamp and slow-motion screw for holding the table, and bringing it to the required position before each individual tooth was cut. This position was determined by the positions of the several divisions under their respective microscopes, due allowance having been made on each micrometer for the tabulated errors of each graduation. The V cutter was then entered a certain depth into the metal at the periphery, and again swung clear. When the "safe" position was indicated by the warning devices the table was revolved to the next division, and the cutter entered again, and so on all round the table without removing the cutter to resharpen it (it had been so prepared as to stand up well under this test of its durability).

In the dividing engine described there are two rows of teeth, one above the other, the upper row spaced 10 minutes apart, and the lower 5 minutes. The purpose of this arrangement was to simplify the driving mechanism and also to provide more equal wear on the bearings of the main spindle of the engine. The tooth-cutting operation was repeated three times round the entire circle in the case of the upper and deeper row of teeth, and twice in the case of the lower row, the final cut in each row being a very light one and effected with a 40° V cutter. The teeth of the cutter were well washed with turpentine during the whole operation, to facilitate the cutting of the metal, which was a very tough alloy of bronze.

The first care after cutting the teeth was to protect them, and for this purpose the rigid circular table provided a permanent support for a substantial cast-bronze shell bolted to it in segments, which completely surrounds the teeth, except where the worm-screws engage them.

§ (9) Driving Mechanism. — It now remained to provide suitable mechanism for

driving the revolving table, and cutting the divisions, each to have a suitable adjustment. Intermittent motion had to be provided for the former, and co-ordinated with this in a very positive manner, a "tracelet," or mechanism operating the actual cutting knife, with its automatic movements. Both are very simple in principle. The intermittent movement of the worm-wheel is arranged as follows: There is, mounted on the worm-spindle A (*Fig.* 2), and rigid with it, a ratchet wheel B, engaged and driven by a pawl C, which revolves on a spindle D. This spindle is separate from, but with its axis in exact prolongation of, the worm-axis. A coarse V-shaped worm E, round which a cord is passed several times, is rigidly mounted on axis D; one end of this cord is attached to a pin on a revolving arm F, the length of which can be varied; the other end of the cord carries a weight. On revolving the arm F the cord round E will be alternately pulled and released. The pawl, therefore, will be revolved round its axis alternately in opposite directions. In one direction it will carry the worm round with it by the pawl engaging in the ratchet wheel, and in the other direction the pawl will slide back over the ratchet wheel teeth. Thus the worm-wheel is alternately revolved and left stationary. The amount of revolution imparted depends on the amount of pull and release given to the cord, or, in other words, to the length of the arm F; this controls the spacing of the graduations. Various adjustments and limiting devices are introduced to make the apparatus as positive in its action as possible.

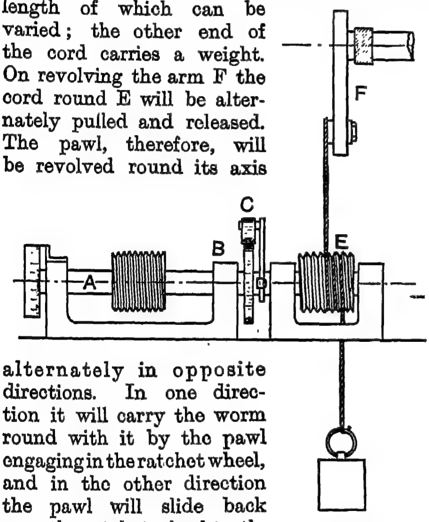


Fig. 2.

As regards the tracelet; two movements of the cutting knife are necessary—one a lowering and lifting movement, and the other a traversing movement; the former to bring the knife into contact with the surface to be graduated and to raise it again, and the latter to cut the division. There are several methods of automatically accomplishing the movements; one of the simplest is shown diagrammatically in *Fig.* 3, where A is the cutting-frame, having at one end a knife or cutter, while the other end is prolonged in order that the cam C, when revolving, may come in contact with it and alternately lift and lower the cutter at the other end. It is important for this cam to

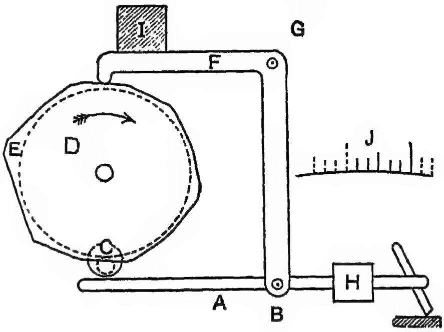


Fig. 3.

revolve at exactly the same rate as the revolving arm F (*Fig.* 2), and to be so adjusted that the cutter is out of contact with the work being graduated while the latter is being rotated by the worm. The shaft carrying the lifting cam C (*Fig.* 3) also carries a pinion geared with a definite ratio (in the case illustrated 1 : 6) into a larger gear wheel D, the spindle of which carries a large cam wheel E, having 6 equally spaced projections round its edge. In close contact with this undulated edge is a "feeler," or a projection of the member F, which is a right-angled frame pivoted at G, into the lower end of which the cutting-frame A is pivoted at B. H is a weight for determining the pressure of the cutting knife on the work, and I is a weight to keep the "feeler" in close contact with the edge of the cam E. The amount of the traversing movement communicated to the cutting frame, and therefore to the cutter, is controlled by the radii of the projections of the cam E.

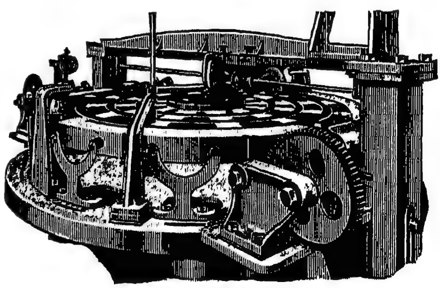


Fig. 4.

In the case illustrated the apparatus would produce graduations somewhat as illustrated at J (*Fig.* 3), the firm lines being cut during one revolution of the cam E. The machine is illustrated in *Fig.* 4.

III. Methods of Reading

§ (10) Method of Reading.—Circles are read by a single pointer, by a vernier, by a reading microscope, or by a micrometer microscope, these being mentioned in their order of increasing accuracy.

§ (11) *The Vernier.*—Suppose a scale divided into equal divisions, with a second scale sliding alongside it, the divisions of the second scale each $1/n$th smaller than those of the main scale. The second scale, or "Vernier," generally contains only n divisions. If the zero of the vernier coincide with a division of the main scale no other division will coincide except the nth, which will coincide with the $(n-1)$th division of the main scale. If the vernier be now moved along a distance of $1/n$th of a division of the main scale, the first division of the vernier, and only the first, will coincide with a division of the main scale; similarly, if it be moved m/n of a main division, the mth division of the vernier will coincide with a division of the main scale. As the eye is very sensitive to lack of continuity in a line, this forms a very sensitive device for splitting the main divisions of a scale. With the aid of a magnifying glass a 6-in. circle can easily be read by a vernier to 30″ of arc, and with a little care to 10″. It will be seen that the divisions of the vernier might equally well be $1/n$th larger than those of the main scale, but in this case the divisions of the vernier must be numbered in the reverse direction from those of the main scale. The device was first described in a treatise by Pierre Vernier, entitled *Construction, usage et propriété du quadrant nouveau de mathématiques*, published at Brussels in 1631.

§ (12) Reading Microscopes.—This device consists of a simple microscope having a fixed cross-hair. This enables a more finely divided circle to be used; but the interval between the cross-hairs must be estimated. If the cross-hair in the microscope be replaced by a scale, the divisions of the main scale can be subdivided by this means, but the final subdivision must still be estimated. The apparent size of the eyepiece scale must obviously agree with the apparent size of the main scale, and this can be adjusted in the same way as the adjustment for run in the case of the micrometer microscope.

§ (13) Micrometer Microscopes.—Here the fixed cross-hair of the reading microscope is replaced by a moving wire, which can be traversed across the field by an accurate micrometer screw, against the pull of a spring. The number of complete revolutions of the screw is counted (if necessary) by a comb cut in a diaphragm in the focal plane. The portions of a revolution are read on a drum attached to the screw. In any case there is a pointer (or zero notch in the comb) which indicates the point being read. When the drum reads zero the cross-hairs should be directly over the pointer or one of the notches in the comb. The run should be adjusted so that the cross-hairs move from one division on the main scale to the next in one or more complete revolutions of the drum. The run can be adjusted by moving the object glass of the microscope toward the circle if the apparent size of the circle division is too small, and away if too great. Any movement of the object glass involves refocussing the microscope. The drum is friction tight on its spindle, and so can be adjusted as necessary; a screw passing through the axis of the milled head used to turn the drum regulates its tightness. The comb or pointer can be moved by turning a screw at the far end of the micrometer box, in prolongation of the micrometer screw (see *Fig.* 5).

In taking a reading the cross-hairs are set over the division on the circle nearest the

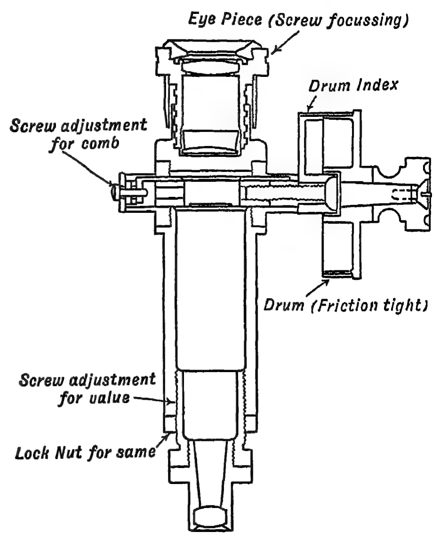


Fig. 5.

pointer. The pointer (or zero of the comb) indicates the division of the circle to be read, and the odd fraction is read off on the drum and comb (if necessary). It is best to have a cross-hair composed of two parallel wires spaced slightly wider than the apparent thickness of a graduation. Two fixed wires at right angles to the movable ones ensure that the readings are always taken on the same part of the graduations. With a 6-in. theodolite it is convenient for the circle to be divided into 10-minute spaces, and for one revolution of the drum to correspond to one division. If the

circumference of the drum be divided into 60 parts, each of these will correspond to 10 seconds, and no comb will be required. With a 12-inch theodolite the main divisions would be 5′, and the microscopes more powerful, so that one revolution of the drum would correspond to 1′. In this case a comb is necessary to count the odd minutes, and one division of the drum corresponds to 1″. In such a case it is convenient to have a lower-powered reading microscope to read the scale direct to the next lowest 5′, only the odd minutes and seconds being read on the micrometer microscopes. The drum reading should be the same whichever graduation be intersected, and it is often convenient to read the graduations on either side of the pointer, as this gives a check on the run of the micrometer, and enables a correction to be made if necessary.

The accuracy of reading depends to a considerable extent on the illumination of the circle. The graduations are engraved lines, and if the light falls from one side the illumination of the two edges differs; the light should fall from above, and in the direction of the length of the lines.

E. O. H.
G. W. W.

Dividing Engines. See "Divided Circles," § (5).

Doppler's Principle: a principle which explains the small change in the wave-length of a line in a spectrum found when the observer or the source of radiation are in motion. By measuring the displacement of such a line, the velocity in the line of sight of the moving source relative to the observer can be determined; other phenomena in spectroscopy can be explained by aid of the principle. See "Spectroscopy, Modern," § (5).

Double-image Prisms, Wollaston's and Rochon's forms of. See "Polarised Light and its Applications," § (12).

Double Refraction. See "Polarised Light and its Applications," § (5). For Fresnel's theory see also § (7).

Drum. See "Sound," § (47).

— E —

Ear, The Human, Description of. See "Sound," § (57) (i.).

Echelette: a type of diffraction grating for work in the infra-red. See "Diffraction Gratings, Theory of," § (12).

Echelon in Spectroscopy: a special form of diffraction grating of very high resolving power. See "Diffraction Gratings, Theory of," § (13).

Efficiency, Effect on Life of (in electric lamps). See "Photometry and Illumination," § (87).

Efficiency of Camera Shutters. See "Shutters, Testing of Photographic," § (3).

Efficiency (of a light source). See "Photometry and Illumination," § (2).

Elasticity of Glass. See "Glass," § (26).

Electron: the fundamental carrier of unit of electric charge. See "Radiology," § (2).

Ellipsoid of Elasticity: the name given by Fresnel to an ellipsoid in a crystal, which determines its optical properties. Its axes are fixed in direction in the crystal. The axes of the ellipse in which a plane wave, passing through the centre of the ellipsoid, cuts the surface, give the directions of vibration in that wave, and their lengths are inversely proportional to the respective velocities of propagation. See "Light, Double Refraction of."

Elliptic Polarisation, Production and Detection of. See "Polarised Light and its Applications," § (15).

Emanations from Radio-elements. See "Radioactivity," § (16).

Embossing, Process of, Glass. See "Glass," § (38).

End Product of Disintegration Series of Radioactive Elements. See "Radioactivity," § (24).

Energy, Partition of, Application of Quantum Theory to. See "Quantum Theory," § (3).

Engraving, Process of, on Glass. See "Glass," § (37).

Enlarger, Photographic: a device for magnifying photographic images. See "Photographic Apparatus," § (10).

Equal Altitude Method for Determining Local Time. See "Surveying and Surveying Instruments," § (26).

Equatorial Mounting for Telescopes. See "Telescope," § (14).

Equivalent Bending Point: a term used in connection with projection apparatus to denote the point in which a ray finally emergent from the apparatus intersects the line of the corresponding incident ray. See "Projection Apparatus," § (4).

Equivalent Bending Surface: a term used to denote the locus of the equivalent bending points for a given symmetrical projector. See "Projection Apparatus," § (4).

Erecting Eyepieces. See "Eyepieces," § (7).

Errors in Observations made in Nautical Astronomy. See "Navigation and Navigational Instruments," §§ (5), (6).

Errors of Sextants. See "Navigation and Navigational Instruments," § (19) (ii.).

Etching, Process of, on Glass. See "Glass," § (38).

Euphonion, B♭: a brass wind-instrument with valves. See "Sound," § (44).

Expansion, Co-efficient of, of Glass. See "Glass," § (27).

EYE, THE

I. Structure

§ (1) The eye, or organ of sight, is probably the most important of the physical sense-organs by which we accomplish the perception of the objective world. A complete treatment of its anatomy and functional processes from the physiological point of view would be outside the purview of this work; but an elementary knowledge of its structure is essential to the comprehension of its properties as an optical instrument, with which the Applied Physicist is primarily concerned, and which is the subject-matter of this article.

In *Fig.* 1[1] a diagrammatic horizontal section of the right eye is shown. The principal features to which we require to call attention here are the following:

The *Eyeball*, which is an elastic body of about one inch diameter, is of the shape shown, being distended at the front to form a spherical cap of smaller radius than the remainder. The wall of the eye consists of three principal layers or coats which, passing from the outside inwards, are met in the following order:

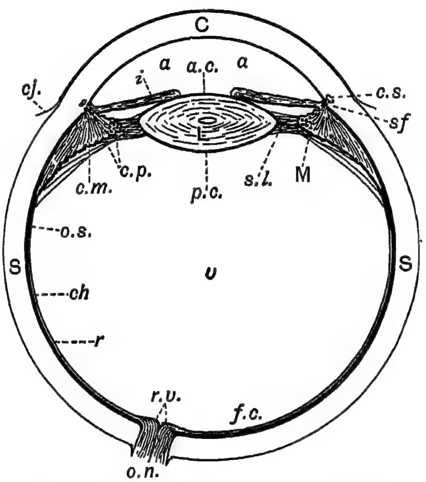


Fig. 1.—Horizontal Section of the Right Eye (Diagrammatic).

aa, aqueous; *v*, vitreous; L, crystalline; C, cornea; S, sclerotic; *ch.*, choroid; *r*, retina; *c.m.*, ciliary muscle; *c.p.*, ciliary processes; M. Müller's ring; *f.c.*, fovea centralis; *i*, iris; *s.l.*, suspensory ligament; *sf*, spaces of Fontana and pectinate ligament; *c.s.*, canal of Schlemm; *a.c.*, anterior capsule of lens; *p.c.*, posterior capsule of lens; *cj.*, conjunctiva; *o.n.*, optic nerve; *o.s.*, ora serrata; *r.v.*, entrance of retinal vessels.

The Sclerotic and Cornea: The *sclerotic* is a white opaque membrane of tough fibrous tissue which extends over about five-sixths of the ball. The *cornea*, which covers the other sixth, is continuous with the *sclerotic*, but is colourless and transparent to admit light to the eye.

The Choroid: This is a layer of highly vascular membrane in close contact with the sclerotic externally and lined internally with dark brown pigment cells.

The Retina: This is the seat of vision. It is a reticulated structure of fibres and cells and is directly connected to the optic nerve. It consists of ten distinct layers, of which the first and second (passing from the choroid inwards) are the receptive layers. The first of these, which adjoins the choroid, is the *hexagonal pigment layer*, a glandular structure in which is secreted a substance known as *visual purple*. That the *visual purple*, which becomes bleached on exposure to light, is associated with vision is regarded as certain, but the part which it plays is unknown, many conflicting theories of its function having been advanced. Next comes the *bacillary layer* or the layer of *rods* and *cones*. These consist of minute bodies, of which the shape is conveyed in their names, and of which vast multitudes are packed side by side perpendicularly to the plane of the retina. The light waves strike these bodies and, by means at present unknown, the radiant energy is converted into some form suitable for transmission along the ramifications of the optic nerve to the brain.

The remaining layers, which need not here be particularised, contain the blood-vessels and the nerve fibres which radiate from the optic nerve to all parts of the layer of rods and cones. It will be observed that the light, before reaching this layer, has to pass through the eight inner layers.

[1] Taken from *Visual Optics and Sight Testing*, by Lionel Laurance, which the reader should consult for fuller information than is here given.

Other books which deal with the physiology of the eye are *Handbuch der physiolog. Optik*, H. von Helmholtz; *Handbuch der Physiologie des Menschen*, W. Nagel; *Optique physiologique*, Tscherning; *Elementary Textbook of Physiology*, T. Huxley.

At the back of the eye, with its centre at the point where the line of sight[1] meets the retina, is the *yellow spot* or *macula lutea*. The spot is slightly elliptical and is about 2·5 mm. in diameter. At its centre is a depression where all the retinal layers, with the exception of the receptive layers, are very much attenuated. This is known as the *fovea centralis*. Its diameter is about 0·25 mm. The acuity of vision is greatest at the fovea, falling off rapidly towards the outer parts of the retina.

The distribution of the rods and cones also varies from the fovea, where there are practically no rods, to the periphery, where there are relatively few cones.

At the point where the optic nerve passes through the wall of the eye the retina is absent and there is no sensitivity to light. This region, which is about 2 to 2·5 mm. in diameter, is known as the *blind spot*.

§ (2) Optical Elements.—The image-producing elements of the eye are the *cornea*, which has already been described, and the *crystalline lens*. The space between the cornea and the lens contains a fluid known as the *Aqueous Humour*, which is practically water with some salts in solution. Its refractive index is nearly that of water and differs little from that of the cornea, which is therefore optically equivalent to a single refracting surface.

Between the lens and the retina is the *Vitreous Humour*, a transparent substance of a consistency similar to the white of an egg (raw). Its refractive index is similar to that of the aqueous humour.

The *Crystalline Lens* is composed of fibres arranged with considerable complexity. It is highly elastic, and its form and position are maintained by a membranous frame known as the *suspensory ligament* which extends from the edges of the lens to the *Ciliary Processes*. These are a direct prolongation of the *Choroid*. Their function is nutritive; they secret a fluid which nourishes the lens and vitreous humour and replenishes the aqueous humour.

Associated with the ciliary processes and the suspensory ligament is the *Ciliary Muscle*. This consists of two parts, the *radiator* and the *sphincter* or *Müller's ring*. Its function is to vary the tension of the suspensory ligament, which results in an alteration of curvature of the posterior surface of the lens, thereby producing the change of power, known as *accommodation*, by which objects at various distances are brought into focus.

The Iris.—This is a diaphragm which limits the aperture of the eye. It consists of a pigmented membrane, in the centre of which is a hole—termed the *pupil*—by which light enters the eye. The iris contains straight radiating fibres, the contraction of which tends to dilate the pupil. This contraction is opposed by a ring of contractile muscle, termed the *sphincter pupillae*, round the edges of the pupil. The operation of the *sphincter pupillae* is involuntary, being due to reflex action. When the retina is subjected to strong illumination, or when a high degree of accommodation is called into play, the *sphincter pupillae* contracts and causes a diminution of the aperture of the pupil.

[1] *Vide infra.*

§ (3) The Eye as an Instrument.—The eye, either alone or in conjunction with external optical instruments, plays an important part in physical experiments. In a great number of determinations one of the factors limiting the accuracy attainable is the precision with which the eye indicates the fulfilment of some particular criterion; for instance, the coincidence of the cross-wires of a telescope with the image of some external object, the equality of brightness or of colour of two illuminated areas, etc. The eye is therefore a physical instrument, of which the capabilities and limitations must be understood equally with those of any other instrument and given due consideration in devising methods and designing apparatus if the experimenter would secure the maximum accuracy in his results.

Unlike most instruments of human design and construction, any one of which is rarely expected to perform more than one special function, the eye has to do a variety of things. These may be grouped under three heads, viz. the perception and measurement of light; the discrimination of colour, and the determination of the position of objects in the field of view. In the following paragraphs we shall study the properties of the eye regarded as an instrument capable of doing these things.

II. The Perception of Light

§ (4) Threshold Phenomena.—The perception of light being fundamental to all the other functions of the eye, we shall first consider the laws governing this perception. There is a minimum quantity of light, known as the *threshold quantity* or *extinction quantity*, which must reach the eye in order that light may be perceived at all. After the eye has become adapted to the dark, a small source of ·02 candle-power, one metre distant, would just be seen when looked at directly. If received indirectly, so that the image falls outside the fovea, a much fainter light could be seen. For most practical purposes it is foveal vision that is important.

In order that a light may be visible it is *necessary* and *sufficient* that the eye should

receive the threshold quantity, *whatever the area of the source*, provided the latter does not subtend a greater angle at the eye than about 50 minutes of arc—or nearly the whole foveal area. This means that for very low intensities the light is completely cumulative, and that, whether concentrated at a point on the retina or spread over an area of 50′ diameter, it produces the same total physio-psychological effect. This is usually expressed by saying that for very feeble light the *visibility*, or apparent brightness, depends only on the total candle-power, and not on the intrinsic brightness of the source so long as its angular size is less than 50′. For larger areas the light required to produce a given brightness near the threshold varies as the linear dimensions of the source. Beyond 4° there is no further cumulative effect, and the apparent brightness is proportional to the intrinsic brightness.

As the brightness increases the area over which complete accumulation takes place diminishes. Thus Paterson and Dudding [1] found the transition from complete to partial accumulation to take place at 10′ for faint sources somewhat above the threshold brightness.

By summing up the results of various observers,[2] S. D. Chalmers [3] arrived at the generalisation that for the true intrinsic brightness of an object to be recognised, *at least the threshold quantity of light must fall on the area of the retina occupied by a single cone; i.e.* an area of 25 to 30 secs. angular diameter.

For very low brightnesses colour is not recognised. From a study of Abney's results for the extinction of colour, Chalmers deduced that the recognition of colour begins at the same stage as the recognition of true brightness, viz. when the threshold quantity of light falls on each cone.

§ (5) Fechner's Law.—At intensities well above the threshold values, but not bright enough to dazzle, the relation between the sensation of light and the stimulus producing it is such that the difference in brightness of two objects which is necessary for one to be recognised as brighter than the other is a constant fraction of the actual brightness: If I is the intensity, and δI_m is the minimum difference which can be detected, $\delta I_m/I$ is constant, $=A$, say. This is known as *Fechner's law*. A, which is about 1 per cent or thereabouts for the average eye, is known as the *Fechner fraction*. The law is only approximate; considerable departures occur at high and low intensities, as already indicated. The value of A, while it may be regarded as constant for practical purposes over the range of medium intensities, in reality follows a shallow curve, having a minimum value when the retinal illumination is about 6·5 metre candles. According to Chalmers [4] the approximate values of A at low intensities are 1·1 per cent at 0·25 metre candles; 1·6 per cent at 0·12 metre candles; 3·2 per cent at 0·03 metre candles. At high intensities A also increases considerably.

[1] *Illuminating Engineer*, May 1915, p. 210.

[2] Aubert (*vide* Tscherning, *Optique physiologique*, p. 211); Abney, *Researches in Colour Vision and the Trichromatic Theory*, 1913, pp. 165, 166, 171, 174, and 183; Paterson and Dudding, *loc. cit.*; Loesser, Ricco, and Charpentier (*vide* Parsons, *Colour Vision*, 120 and 118); Piper, *Z. f. Psychol. u. Physiol. d. Sinnesorg*, 1903, xxiii. 98.

[3] *Trans. Opt. Soc.*, 1919, xx. 297.

III. Colour Phenomena

§ (6) Visibility.—We need not deal here with the theory of the mechanism by which the wave-length or vibration frequency of the light imparts to the sensation that character which we know as colour; but some of the properties of colour vision are of great practical importance.

If we look at two areas illuminated by light of different wave-lengths and vary their relative intensities, we perceive that there is a certain condition in which we appraise them as equally bright, although their colours may be quite different. The criterion of equal brightness in such a case is probably purely psychological, since there is nothing in the external stimuli, nor, as far as is known, in their physiological effects, which can be said to be equal in any ordinary sense of the word. The mind appreciates the illuminations as equally bright when they make equal claims on the attention. What determines this claim is obscure, and in any case does not concern us here: the important point is the experimental fact that for radiations of different wave-lengths there is a unique relation between the quantities received by unit area of the retina for which the radiations will be regarded as equally bright.

The degree of brightness produced by a given quantity of energy falling on unit area of the retina in unit time depends on the wave-length of the radiation. For blue and yellow-green lights, for instance, to appear equally bright, the energy flux from the blue must be many times as great as from the yellow-green. The blue radiation is said to be of lower *visibility* than the yellow-green. The inverse ratio of the energy fluxes required to give equality of brightness may be regarded as the *relative visibility* of the two colours; or, if we take the visibility of the yellow-green as unity, the inverse ratio of the fluxes measures the *visibility* of the blue.

The visibility of monochromatic illumination throughout the visible spectrum has been

[4] *Loc. cit.*

determined by various observers.[1] For a normal eye the curve connecting visibility and wave-length is of the form shown in *Fig.* 2, maximum visibility occurring for a wave-length about $0{\cdot}55\mu$.

The curve can be represented to a close approximation by the following expression:[2]

$$V_\lambda = A\left(\frac{R_1}{\lambda}e^{1-\frac{R_1}{\lambda}}\right)^\alpha + B\left(\frac{R_2}{\lambda}e^{1-\frac{R_2}{\lambda}}\right)^\beta + C\left(\frac{R_3}{\lambda}e^{1-\frac{R_3}{\lambda}}\right)^\gamma,$$

where

$A=0{\cdot}999$	$R_1=0{\cdot}556$	$\alpha=200$
$B=0{\cdot}04$	$R_2=0{\cdot}465$	$\beta=400$
$C=0{\cdot}095$	$R_3=0{\cdot}610$	$\gamma=1000$

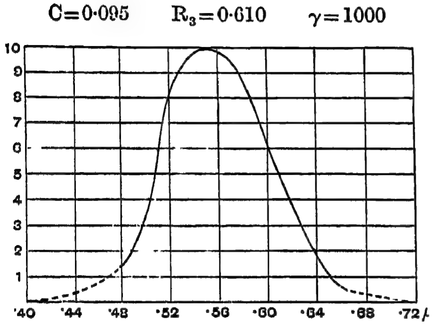


Fig. 2.

§ (7) White Light.—A stimulus in which components of all wave-lengths were present in brightnesses proportional to the ordinates of the visibility curve would not produce white light; for in none of the sources of radiation which the eye regards as white is the energy distributed uniformly in the spectrum. For any actual source, the relative quantity of light sensation produced by each wave-length is proportional to the product of the ordinates of the visibility curve and the energy distribution curve of the source for the wave-length in question. This product is termed the *luminosity* of the source for this wave-length. If the luminosity for each wave-length be plotted, a new curve will be obtained differing from the visibility curve to the extent that the energy distribution curve of the source departs from a horizontal straight line. The energy distribution curve of a "black body" at 5000° C., which can be calculated from Wien's law, may be taken as the standard "white" light. In practice the radiation from the crater of an electric arc is about the whitest artificial light, and is the source to which the majority of the colour vision determinations of Abney and others refer.

§ (8) Hue and Saturation.—In the case of light which is not white, the eye appreciates three factors—*Hue, Degree of Saturation,* and *Brightness.* As we shall see later, any colour, with the exception of purples and magentas, can be produced by mixing a proper proportion of white light with monochromatic radiation of suitable wave-length. The *hue* of the colour is specified by the wave-length of the monochromatic radiation; the degree of *saturation,* or rather lack of saturation, by the percentage of white in the total. The perception of hue is, of course, entirely subjective, and the rate at which it changes with wave-length is very irregular. If two adjacent areas be illuminated with monochromatic light of different wave-lengths, the extent by which these must differ for a hue difference to be perceptible can be determined at all parts of the spectrum. This difference has been measured by various experimenters.[3] The curve obtained by Jones is shown in *Fig.* 3. The inverse of this curve

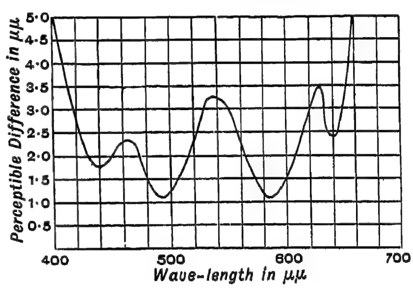


Fig. 3.

indicates the sensitivity of the eye to variation of wave-length at different parts of the spectrum.

A slight change in hue of the light reflected or transmitted by a substance is frequently the fiducial criterion in physical and chemical processes, and its detection with the utmost sensitivity is therefore of importance. For instance, the "dose" of X-rays administered for the treatment of ringworm, etc., is measured by the change of hue of Saburaud pastilles, which change from lemon to orange with the

[1] Koenig, "Helmholtz Festgrüss," *Beit. z. Psych. u. Physiol. d. Sinnesorg.*, Hamburg u. Leipzig, 1891, p. 309; also *Verh. d. Physik., Ges. Berl.* 1892, ii. 10; Thürmel, *Ann. d. Physik*, 1910, xxxiii. 1154; Nutting, *Bur. of Standards Bull.*, 1908, v. 261; 1911, vii. 235; *Trans. Illum. Eng. Soc.* (U.S.), 1914, ix. 633; P. Reeves, *ibid.*, 1918, xiii. 108; Ives, *Phil. Mag.*, 1912, pp. 149, 352, 744, 845, 853; Hyde and Forsyth, *Astrophys. Journ.*, 1915, xlii. 285; Hartman, *ibid.*, 1918, xlvii. 83; Hyde, Forsyth and Cady, *ibid.*, 1918, xlviii. 65; Coblentz and Emerson, *Bur. Stds. Sci. Papers*, 303, and *Bull. Bur. Stds.*, 1918, xiv. 167. This last paper contains a very complete bibliography of papers on visibility.

[2] H. E. Ives, *Journ. Franklin Inst.*, 1915, clxxx. 409.

[3] Steindler, O., *Ak. Wiss., Wien*, Jan. 1906, 115, 2A; L. A. Jones, *Journal Opt. Soc. Amer.*, 1917, i. 63; *vide* also P. G. Nutting, *Bull. Bur. Stds.*, 1909–1910, vi. 89.

proper exposure. The accuracy with which the latter tint can be matched with a standard determines the accuracy with which the dose can be administered.

The sensitivity of any colour criterion will be greatest if the final colour is near one of the minima of *Fig.* 3. In arranging such reactions an attempt should always be made to bring the tint to a sensitive region by the addition of colouring-matter which is inert to the reaction, or, possibly, by the use of suitable colour-filters. In using such expedients, however, care must be taken to avoid suppressing those wave-lengths which are affected by the reaction.

§ (9) The Trichromatic Basis of Colour.—As an expression of what the eye perceives, the specification of colour by its hue and degree of saturation is the most natural; but it does not give the readiest means of co-ordinating colour-mixture phenomena with the properties of the receiving mechanism. It is found that if monochromatic light of three suitable wave-lengths—a red, a green, and a blue—be mixed in suitable proportions, white light will be produced. Further, it is possible to match the *hue* of any colour whatever by mixing some pair of these primaries in suitable proportions. In general the saturation of this mixture will differ from that of the colour. If it is too highly saturated an exact match can be obtained by adding white, or, alternatively, by adding suitable amounts of all three primaries in the proportions which make white. That is to say we can match such a colour both in hue and saturation by a suitable mixture of our three primary colours.

It is necessary to attach a meaning to the term "quantity" when we refer to mixtures of certain relative quantities of light of different colours. It is found to be most convenient in practice to choose the units for the three primaries such that an equal number of units of each will, when mixed, give white light. It should be noted that this is quite an arbitrary choice; there is no physical, physiological or psychological property of the radiations which are equal for equal quantities as defined on this basis. For instance, they would not appear equally bright if separated. It is, however, a convenient system for expressing the numerical data of colour mixing.

We may therefore express any colour, of which the saturation is not too high, in terms of three primaries R, G, and B, by three quantity coefficients a, a', and a'', which give the numbers of units of each primary, as above defined, which are required to match the colour, thus

$$\mathrm{C}=a\mathrm{R}+a'\mathrm{G}+a''\mathrm{B}.$$

We may find that the colour to be matched is of greater saturation than the two-primary mixture which gives its hue. In this case an exact match is impossible, since addition of the third primary can only produce mixtures of still lower saturation. We may, however, add white to the colour, thereby reducing its saturation to such an extent that it can be matched by combination of the two appropriate primaries. Expressing this white by its equivalent in terms of the three primaries we have the relation

$$\mathrm{C}+\alpha\mathrm{R}+\alpha\mathrm{G}+\alpha\mathrm{B}=a\mathrm{R}+a'\mathrm{G},\text{ say,}$$

or

$$\begin{aligned}\mathrm{C}&=(a-\alpha)\mathrm{R}+(a'-\alpha)\mathrm{G}-\alpha\mathrm{B}\\&=b\mathrm{R}+b'\mathrm{G}+b''\mathrm{B},\text{ where }b''=-\alpha.\end{aligned}$$

Thus, although we cannot match such a colour by any actual mixture of our three primaries, we can arrange a slightly different colour match from which we are able to *specify* our original colour in terms of the three primaries by introducing a negative coefficient for some one of them.

It is evident that instead of adding white to the colour we could have adopted the alternative procedure of adding the third primary alone, and have obtained the colour match

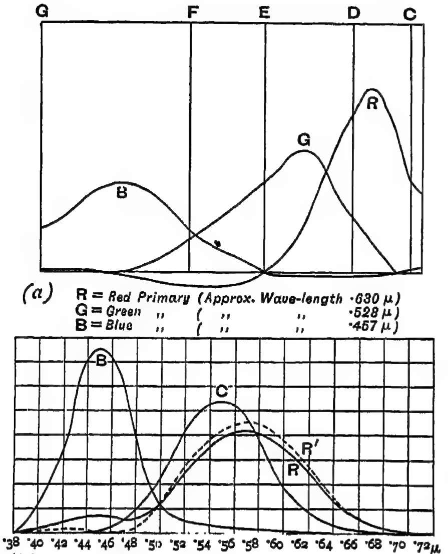
$$\mathrm{C}+\alpha\mathrm{B}=b\mathrm{R}+b'\mathrm{G},$$

the final specification for C being the same in each case. It depends on the construction of the colour-matching apparatus employed whether it is more convenient to add white or one of the primaries.

There is a comparatively wide range of wave-length within which each of the three primary colours may be chosen to fulfil the fundamental condition of being miscible with the other two to make white. For colour-matching purposes, however, it is undesirable to choose any primary near an extremity of the possible range, as in such a case some of the mixture colours are of very low saturation.

§ (10) Mixture Curves.—If we plot at each wave-length the quantities of the three primaries required to match the spectrum both in colour and brightness we obtain a set of three curves, which are termed the *Mixture Curves*. The curves for the primaries used in Clerk Maxwell's colour-mixing apparatus are shown in *Fig.* 4 (*a*). The heights at which the ordinate at any wave-length cuts the three curves R, G and B give the relative quantities of the primaries required to match the spectrum colour. For any one primary the aggregate quantity required to match every part of the spectrum is clearly $\int_0^\infty s_\lambda d\lambda$, where s_λ is the quantity of that primary required at the wave-length λ. The value of the integral is measured by the area enclosed between the curve and the base-line, parts below the base being regarded as negative.

Since the total effect of the spectrum is to give white light, and since our units are chosen so that equal quantities of the primaries



(b) FIG. 4.—The Fundamental Sensation Curves of the Spectrum.

Full lines mean values of Koenig's five observers; dashed line, new red curve, obtained by different choice of blue primary dictated by luminosity considerations.

make white, the areas of the three mixture curves are equal.

§ (11) YOUNG'S THEORY OF COLOUR VISION.[1] —The fact that all colours can be reproduced for practical purposes by mixing suitable quantities of three primary colours suggested to Thomas Young that there are three receiving mechanisms in the eye which when stimulated give rise to primary colour sensations of red, green, and blue respectively. Each sensation is stimulated by radiation from a considerable range of the spectrum, though for each there is a maximum response at a certain wave-length. The receiving mechanisms have, in fact, the properties of damped resonators with natural periods corresponding to the wave-length which produces maximum response. The resultant colour effect produced by radiation of any wave-length is the mixed sensation due to the partial stimulation of the three primary sensations.

If we had three colours, each of which stimulated one *and one only* of the primary sensations, these would form the natural system of primaries to which all colour mixture phenomena should be referred; since, by mixtures of such primaries, it would be possible to reproduce any possible colour, both in hue and purity, without introducing negative coefficients.

Unfortunately we cannot find three such colours. If we choose a red of wave-length about $0·7\mu$ or longer, the resulting colour is almost entirely due to pure red sensation; and if we take a blue from a very short region about $·446\mu$ we will, according to the most recent work,[2] obtain nearly pure blue sensation; but there is no colour which stimulates the green sensation alone, every green, to the normal eye, stimulating not only green sensation but varying amounts of the other two. There is, however, a point in the spectrum, about $0·50\mu$, where the red and blue sensations are equally stimulated. The colour here may be regarded as made up of equal stimulations of red, blue, and green, plus the remainder of the green; that is to say, it is green diluted with a certain amount of white. Thus at $0·50\mu$ the spectrum is the same *hue* as the pure green sensation would be, were the others absent or atrophied, *but it is less saturated.*

By making colour matches with these primaries we can construct their mixture curves as described above. By combining these with data for colour-blind observers of different types[3] it is possible to construct the *sensation* mixture curves, *i.e.* curves showing the amount of stimulation of the three primary sensations at different wave-lengths. In *Fig.* 4 (*b*) are shown the sensation curves as obtained by Koenig and modified by H. E. Ives.[4]

Every colour can be specified in terms of these three sensations by coefficients indicating the degree to which they are stimulated; thus

$$C = aR + a'G + a''B.$$

The sum of the coefficients $a + a' + a''$ is termed the *quantity* of the colour, being the sum of the quantities of the three sensations.

§ (12) COLOUR GEOMETRY.—Since colours can be specified by three coefficients in this way it is evident that colour mixture data can be represented by points in a three-dimensional figure, and that geometrical methods can be used to solve colour mixture problems. A full treatment of mixture diagrams would be beyond the scope of this article,[5] but reference may be made to the colour triangle. This is a section of the three-dimensional diagram (the colour pyramid) by a plane which makes equal intercepts with the axes.

In the colour triangle the three primaries are represented by the corners of an equilateral

[1] In its later form, as modified by v. Helmholtz, this is known as the Young-Helmholtz Theory of Trichromatic Vision.

[2] H. E. Ives, *Journ. Franklin Inst.*, 1915, clxxx. 409.

[3] See, for instance, Abney, *Colour Vision*, chap. vii.

[4] *Journ. Franklin Inst.*, 1915, clxxx. 409.

[5] *Loc. cit.*

triangle. Any point in the figure represents a colour $aR + a'G + a''B$, where a, a', and a'' are proportional to the perpendicular distances from the point in question to the sides of the triangle opposite R, G, and B respectively. A triangle can be constructed for any set of primaries used for colour-mixing; but a particular colour will occupy different positions in different triangles, since different proportions of the primaries are required to match it. White, however, will always be represented by the centroid of any triangle; for, *by choice of units*, equal quantities of the primaries make white.

Any possible hue obtainable by mixing two colours is represented by a point on the line joining their positions in the colour triangle, dividing the line in the inverse ratio of the quantities of the two in the mixture.

Since the sum of the perpendiculars from any point to the three sides of the colour triangle are equal, all points on the diagram represent colours in equal quantities, *i.e.* equal values for $a + a' + a''$. We may take this as unity. Any point, therefore, represents unit quantity of colour. It then follows from simple geometry that the colour obtained by mixing two colours represented by points A and B, in proportions x of A to y of B, will be given by the point which divides the line AB in the ratio $y : x$. If the line joining any two colours passes through white, the colours are complementary, since by mixing them in suitable proportions white is produced.

For many purposes the relative luminosities of the constituent colours in a mixture are required. This is not equal to the relative quantities of the colours as obtained from the colour triangle, since equal quantities of different colours on the arbitrary quantity convention on which the triangle is based are not of equal luminosity. We may obtain the relative luminosity of unit quantities of any colour if we know the relative luminosities of equal quantities of the three primaries. Suppose we have unit quantity of colour, C, given by $C = aR + a'G + a''B$, where $a + a' + a'' = 1$. Its luminosity, L_C, is clearly $aL_R + a'L_G + a''L_B$, where L_R, L_G, and L_B are the luminosities of unit quantities of the three primaries. L_R, L_G, and L_B are easily determined, so that it is possible to calculate the relative luminosity of unit quantity of any colour of which the trichromatic coefficients are known.

Having found from the triangle the relative quantities of any two colours C and C_1 in a mixture, it is only necessary to multiply the result by the ratio L_C/L_{C_1} to obtain the relative luminosities of the constituents. Conversely, if it is desired to determine the colour which results from mixing, in a given luminosity ratio, two constituents C and C_1 of which the colour triangle coefficients are a, a', a'' and a_1, a_1', and a_1'', respectively, we first calculate $L_C (= aL_R + a'L_G + a''L_B)$ and $L_{C_1} (= a_1 L_R + a_1' L_G + a_1'' L_B)$, the relative luminosities of unit quantities of the two colours, from which we can find the quantity ratio corresponding to the required luminosity ratio. The mixture colour is then found in the triangle by dividing the line joining the constituents inversely as the quantity ratio.

Since the ratios of L_R, L_G, and L_B to one another differ for different sets of primaries, the relative luminosities of unit quantities of different colours also vary. That is to say, the relation between quantity and luminosity differs for different colour triangles. The final result of any colour calculation is, however, the same on whatever triangle the specifications are based.

Ives[1] has calculated that the relative luminosities to be attributed to the sensation primaries are $L_R = 0{\cdot}648$; $L_G = 0{\cdot}336$; $L_B = 0{\cdot}016$. These are expressed so that their sum is equal to unity. The luminosity of unit quantity of white, for which $a = a' = a'' = \frac{1}{3}$, is therefore $\frac{1}{3}(L_R + L_G + L_B) = \frac{1}{3}$.

It is necessary for many colour measurements to have the spectrum colours plotted in the triangle. This is done by taking the values a, a', and a'' for different wave-lengths from the mixture curves of the spectrum, *obtained with the primaries for which the triangle is constructed*, and plotting the points at their proportionate distances from the three sides. If as primaries we take the three *sensations* we use the sensation curves of *Fig.* 4 (*b*). In *Fig.* 5 the spectrum is shown in the fundamental colour sensation triangle, the new red curve of *Fig.* 4 (*b*) being employed in its construction.

We can now express any colour, of which the sensation coefficients are known, in terms of its spectral hue and degree of saturation. Take the colour whose sensation values are $\cdot55R + \cdot30G + \cdot15B$. It is represented by P, *Fig.* 5. The line WP produced cuts the spectrum curve at H, where the wave-length is $\cdot611\mu$.

The hue of the colour is that of the spectrum at $\cdot611\mu$, and it is diluted with white to the extent of 100 HP/HW per cent $= 43{\cdot}3$ per cent on the quantity basis.

To obtain the percentage *luminosity* of the white we must multiply this figure by the ratio of unit quantity of white to unit quantity of the total colour. As we have already seen

$L_W = 0{\cdot}333$.

$$L_C = 0{\cdot}55 \times 0{\cdot}648 + 0{\cdot}30 \times 0{\cdot}336 + 0{\cdot}15 \times 0{\cdot}016 = 0{\cdot}459.$$

[1] *Journ. Franklin Inst.*, 1915, clxxx. 409.

The percentage luminosity of white is therefore

$$43 \cdot 3 \times \frac{0 \cdot 333}{0 \cdot 459} = 31 \cdot 4.$$

The colour triangle readily gives other information about the colour properties of the eye. If a straight line drawn through W cuts the spectrum in two points, the colours at these points will be complementary, producing white when mixed in inverse proportion to their distances from W. The colours of the spectrum from about $\cdot 493\mu$ to $\cdot 56\mu$ have no complementaries in the spectrum. They are complementary to colours lying in the region approximately bounded by RWB, the region of purples and magentas. We see now why purples and magentas are exceptions to the rule that any colour can be made up of monochromatic light plus white; in this region a line joining white to the colour does not intersect the spectrum. A colour of this kind has to be specified, on the hue and purity scale, by the *amount of its complementary spectral hue which must be mixed with it to make white.*

Precisely the same method of use applies to colour triangles constructed for actual primary colours instead of the fundamental sensations. The position of the spectrum in the diagram will be different; the three wave-lengths chosen as primaries will occupy the corners, while the location of the other wave-lengths will depend on the form of the colour mixture curves for the primaries concerned.

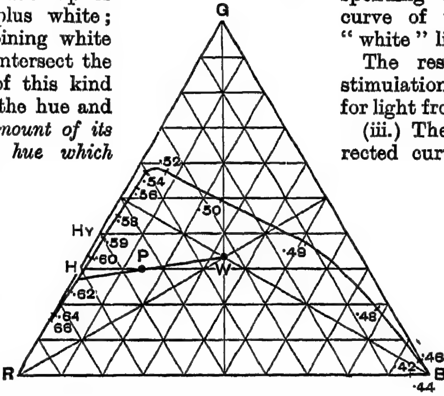


FIG. 5.—The Fundamental Colour Sensation Triangle.

In practical colorimetry actual spectrum primaries must, of course, be used. These may either be isolated spectroscopically or may simply be obtained by coloured glasses. For the co-ordination of colorimetric data it is necessary to reduce the data obtained with the mixing instrument to a standard system; and the most suitable is that based on the three fundamental sensations. The methods of reduction are outside our present scope. They are fully developed by H. E. Ives in the paper already quoted,[1] in which the parallel problem of the reduction of measurements on the hue and purity system to the trichromatic system and *vice versa* is also treated.

§ (13) DOMINANT HUE OF FILTERS.—As an illustration of the application of mixture curves and colour triangles, a case that often arises in the laboratory may be quoted. It is frequently desired for various purposes to use a "filter" of coloured glass or other material to render light approximately monochromatic. If a knowledge of the wave-length of the dominant hue is required—as, for instance, in using such filters for optical pyrometry—the following process has to be followed:

(i.) The spectrophotometric curve for the filter is obtained.

This gives the percentage of the incident light transmitted at all wave-lengths.

(ii.) If the energy distribution in the spectrum of the light with which the filter is to be used differs from that of the "white" light to which the sensation mixture curves refer, it is necessary to multiply the ordinates of these curves by the ratio of the corresponding ordinates of the energy curve of the source and of the "white" light.

The resulting curves give the stimulation of the three sensations for light from the source in question.

(iii.) The ordinates of these corrected curves must be multiplied by the transmission coefficients of the filter.

The double reduction of (ii.) and (iii.) may be done in one process since the intermediate curves are of no special interest.

The ratios of the areas of the final mixture curves to their initial values give the coefficients a, a', and a'', which express the colour of the transmitted light in terms of the three sensations. This can be plotted in the sensation triangle, *Fig.* 5, and the predominant wave-length and degree of saturation obtained as described in the preceding paragraph.

§ (14) PHENOMENA AT LOW INTENSITIES.—In the foregoing paragraphs the visibility curve and the sensation mixture curves have been regarded as invariant. For medium intensities this is practically the case; but at low intensities the colour perception is not represented by the curves as given in *Figs.* 2 and 4.

As the intensity of the spectrum is reduced, the maximum of the visibility curve shifts towards the blue, the eye becoming relatively less sensitive to the red end of the spectrum. At low intensities the red disappears entirely; the orange becomes yellow, and the green bluish, this process continuing until nothing is left but a faint bluish grey. With still further reduction, the last trace of colour

[1] *Journ. Franklin Inst.*, 1915, clxxx. 409.

isappears, leaving only a nondescript grey. 'his variation of colour sensitiveness with ıtensity is known as the *Purkinje effect.* Jnless care is taken to work at suitable ıtensities, it may introduce undesirable omplications into colorimetric measurements.

Not only do the colour sensations vary with ntensity, but they are not constant for all ›arts of the retina. Within the *macula lutea* ›r "yellow spot" there is considerable absorpion towards the blue end of the spectrum.[1] Chis region is therefore relatively more sensiive to red than regions just outside it. As ve recede from the centre the visibility curve changes in approximately the same way as when we reduce the intensity, the peripheral portions of the retina being quite colour-blind, and all colours appearing of the nondescript grey already referred to. The reason why this is never noticed in ordinary vision is because of the very great difficulty of seeing anything distinctly at any distance from the fovea.

To explain these phenomena the "Duplicitats Theory" of von Kries assumes that the colour sensitive mechanisms are associated with the cones, and that the colourless sensation, which predominates at low intensities and towards the periphery of the retina, is located in the rods. The cones are only excited at comparatively high intensities, and the Purkinje effect is explained by the gradual passage from cone to rod vision as the intensity is reduced. The relative concentration of the cones is greatest at the centre of the retina, diminishing practically to zero towards the outskirts. Hence the passage from chromatic to colourless vision from the centre outwards. In most eyes there are no rods at all at the centre of the fovea, and in this region the Purkinje effect is found to be absent, there being no light visible when the colour has disappeared.

§ (15) Dark Adaptation.—The sensitiveness to faint light depends very considerably on the length of time which has elapsed since last the retina was exposed to a bright illumination. Every one is familiar with this phenomenon. It requires about ten minutes to become completely dark-adapted, and the sensitivity to feeble illumination is then about a hundred times as great as it is immediately after exposure to daylight.[2] According to Parinaud,[3] dark adaptation is entirely due to the increased sensitivity of the colourless sensation. Thus at the red end of the spectrum, where there is no change of hue and no passage into colourless vision on diminishing the intensity, there is no gain in sensitiveness due to adaptation. The effect of adaptation commences about the C line ($\lambda = \cdot 656\mu$) and becomes of increasing importance towards the blue end. *Fig.* 6 shows Parinaud's results. The lettering indicates the Fraunhofer lines. The ordinates indicate the quantity of light necessary in order to be perceived. The upper curve refers to the dark-adapted eye and the lower to the non-adapted eye. The unit chosen is the quantity required by the adapted eye in the neighbourhood of E ($\cdot 527\mu$, green). The figure shows that the non-adapted eye requires 100 units[4] at this wave-length as against 1 for the adapted eye, and 1500 units for blue light in the neighbourhood of the G line, as against 100 for the adapted eye. The increase of sensitivity on adaptation is not due, however, to increased sensitivity of the colour perception, but only of the colourless rod vision. At the centre of the fovea, where, as we have seen, there are no rods, and in consequence no Purkinje effect, there is also, according to Parinaud, no dark adaptation.

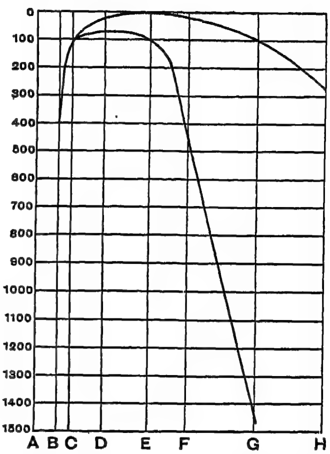


Fig. 6.—Position in Solar Spectrum.

§ (16) Visual Fatigue and Successive Contrast.—When a bright object is viewed steadily for some time it becomes appreciably fainter, due to the elements of the retina on which the image falls becoming fatigued. The effect of fatigue remains for some time after the stimulus ceases; and, if the gaze be transferred to a uniformly illuminated surface, a dark "negative image" of the bright object will be seen, due to the diminished sensitivity of the part of the retina previously occupied by it. If the object is coloured, that sensation which is most stimulated is also most fatigued, and the negative image is of a colour approximately complementary to that of the object.

[1] Abney, *Colour Vision*, p. 91.
[2] Abney, *loc. cit.* p. 120.
[3] *Ann. d'Oc.*, 1894, t. cxii. 228. *Vide* Tscherning, *Optique physiologique*, p. 226.
[4] Compare Abney's result *supra*.

This phenomenon is known as *Successive Contrast*. It is not, of course, a true contrast effect. True or simultaneous contrast is observed when two areas of different brightness adjoin each other. Near the junction the brighter area appears even brighter and the darker area even darker than at places remote from the junction. If the areas are of slightly different colours the difference is intensified by contrast in the immediate neighbourhood of the junction. Contrast effects of this character disappear almost entirely if the areas are separated by a black line. Diffuseness of the border appears to favour the effect.

If a screen is illuminated by two lights, one of which is white and the other coloured, and an object is placed in front of the screen, it will cast two shadows, one of which is illuminated solely by the white light, and the other solely by the coloured light. The former, which ought to appear white, appears coloured. The colour is approximately complementary to the colour of the coloured light, though according to Abney it is not always exactly complementary. Tscherning [1] accounts for contrast colours obtained in such circumstances as due to defective judgment of white. There is always a tendency to take a large illuminated area as white, unless it is very strongly coloured. Even then it appears whiter than it should, the "white" with which it is mentally compared being tinted with the colour in question. If, then, a small area of true white exists in such a background, it is poorer than the temporary standard white in the colour with which the latter is adulterated, and, therefore, appears of the complementary colour.

Contrast effects, both of brightness and colour, are of great importance in connection with the design of photometers and colorimetric apparatus.

§ (17) Time Effects.—When a light is suddenly turned on it is not immediately perceived, nor, when it is turned off, does perception immediately cease. The time lag in the second case exceeds that in the first case, with the result that vision of an instantaneously exposed light (for instance, an electric spark) persists for an appreciable time. Thus if a light is rapidly intermittent, the impression of one exposure may still persist until the next exposure, and the light will be seen continuously. The period which may elapse between successive exposures without the appearance of flicker may be termed the duration period. The total duration, of course, exceeds this considerably. E. L. Nichols [2] and E. S. Ferry [3] have determined the duration period as defined above. The general results are as follows:

1. The persistence decreases as the intensity increases.
2. It varies with wave-length, being greater towards the ends of the spectrum than at the middle.
3. It is greatest for very short exposures.
4. It differs for different eyes.
5. An average duration under ordinary conditions is from $\frac{1}{30}$ to $\frac{1}{10}$ second.

§ (18) Talbot's Law.—The apparent intensity of an intermittent light is diminished, even if the intermittence is so rapid that continuous vision results. Talbot [4] enunciated the law that the apparent intensity of the light in those circumstances bears the same ratio to its actual intensity as the time of exposure to the total time. Thus if the period of exposure is equal to the period of eclipse the intensity is diminished 50 per cent. Talbot's Law is utilised in the employment of rotating sector discs to cut down the intensity of light without altering its spectral composition. Contrary to the results of earlier workers,[5] E. P. Hyde [6] has shown that the law holds for all sector openings, at any rate down to 10°, and for all colours.

§ (19) Visual Diffusivity.—The time lag which precedes the perception of light has not so far been measured; but it is easy to measure its variation under different conditions. H. E. Ives [7] has shown that it diminishes as the intensity increases, and increases from the red end of the spectrum to the blue. It is greatest of all for the colourless rod vision, this, in fact, constituting the "after image" which is seen about half a second after a bright flash. Ives shows that these various phenomena are all consistent with the theory that the transmission of impressions from the retina to the brain is in accordance with the physical laws of conduction, the stimulus being transmitted through matter having a coefficient of *diffusivity* which varies with wave-length and intensity. He deduces that for coloured light the diffusivity is a rectilinear function of the logarithm of the stimulus intensity; and the time lag, t, which varies inversely as the diffusivity, $=1/(a \log \mathrm{I}+b)$, a and b being constants for the wave-length concerned. Thus the difference of lag for two different intensities is

$$t_1-t_2=\frac{1}{a \log \mathrm{I}_1+b}-\frac{1}{a \log \mathrm{I}_2+b}.$$

[1] *Optique physiologique*, p. 221.
[2] *Am. J. Sci.*, 1884, xxviii. 243.
[3] *Ibid.*, 1892, xliv. 192.

[4] *Phil. Mag.*, 1834, v. 327.
[5] Ferry, *Phys. Rev.*, 1893, i. 338; see also Lummer and Brodhun, *Zeits. Instrumentenk.*, 1896, p. 229, for a general discussion of Talbot's Law.
[6] *Bull. Bur. Stds.*, 1906, ii. 1.
[7] *Phil. Mag.*, 1917, xxxiii. 18.

For two different colours, red and blue say, at the same intensity,

$$t_B - t_R = \frac{1}{c \log \overline{I+d}} - \frac{1}{a \log I+b}.$$

In a series of papers[1] Ives and Kingsbury have shown the importance of visual diffusivity in the theory of the Flicker Photometer.

Many interesting phenomena related to the subject-matter of this and preceding paragraphs have been described by Shelford Bidwell,[2] but a description would be outside our present scope.

IV. Dioptric Properties

§ (20) Accommodation.—The power of the normal eye in repose is approximately 59 diopters. By the process of *Accommodation* the power can be varied at will so as to bring objects at various distances into sharp focus. This power is greatest in children and diminishes regularly with age. Its range is about 10 diopters at 20 to 25, about 2 diopters at 50, and is negligible above 60 years of age. While these figures apply approximately to the conditions of daily use, the actual amount of accommodation called into play is not equal for all parts of the eye, the accommodation at the centre of the pupil being greater than near the periphery. This is due to the fact[3] that during accommodation the anterior surface of the crystalline lens increases in curvature near the centre but actually flattens near the periphery. Nature compensates for this by reducing the size of the pupil when the eye is accommodated for near objects, thereby cutting out the peripheral regions, which would otherwise spoil the definition. If the pupil is dilated by some drug which does not affect the range of accommodation,[4] the deterioration of the definition for near objects is usually marked.

§ (21) The Pupil.—This is the name given to the circular aperture of the eye. It is usually concentric with the optic axis, *i.e.* with the line through the centres of curvature of the various refracting surfaces. The centring of the eye is never exact, though the departures from it are rarely important. The most common defect is that the centre of curvature of the cornea is not on the axis of the crystalline lens. The error may sometimes amount to a quarter millimetre. Occasionally the pupil is displaced outward from the axis. Relatively large defects of centring do not appear to detract from the acuity of vision.

The pupil diameter alters under various influences. The mechanism by which this happens is complex and is not completely understood. As mentioned in the last paragraph it varies with the degree of accommodation and also with the intensity of the light. As regards the former effect, J. W. French[5] found it to be negligible until, for a normal eye, the distance of vision came within about 30 cm. Thereafter the contraction was rapid. In a particular experiment the pupil shrank from 6·5 mm. at 30 cm. to 3·5 mm. at 10 cm.

As regards the second effect French found that the pupil area could be represented over a large range of brightness by the relation

$$A \propto I^{-\frac{1}{5}}.$$

Thus, in order to reduce the area of the pupil by half, the intensity would have to increase thirty-two times. The above equation is for the central area covered by the *macula lutea.* French also investigated the effect on the pupil of light falling on other zones of the retina. The most sensitive zone is the one just outside the macula, the index for this zone being $-\frac{1}{4}$. On going further from the centre the effect diminishes, the peripheral zone being very insensitive. In French's words: "This portion of the retina requires all the light it can get, and the pupil opens out to nearly its maximum diameter and responds but little to variations of intensity."

It is usually stated that the two pupils always vary together, even if the intensity to which they are exposed is widely different. This view is undoubtedly based on considerable evidence under ordinary conditions of vision; but French describes experiments in which one pupil remained of constant size under constant illumination, while the other varied over a wide range under varying illumination.

This independence may not be common to all eyes, but it is evidently unsafe to assume that in all circumstances the two pupils will be of equal diameter.

§ (22) The Line of Sight: Fixation.—When the observer "looks at" a particular object he is said to *fix* it. The line joining the front nodal point of the eye to the *point of fixation* is termed the *line of sight.* The image of the point of fixation is formed on the fovea centralis, but only a very minute area is fixed at a time. In fact, however close two points may be, provided they can be seen as separate points, it is also possible to say that one of them is being looked at rather than the other.

Contrary to what might be expected, the line of sight is not coincident with the optic axis. It is inclined inwards and downwards with respect to it by 5 to 7 degrees. Since it passes

[1] Ives and Kingsbury, *Phil. Mag.*, 1914, p. 708; *Phil. Mag.*, 1916, p. 290; Ives, *Phil. Mag.*, 1917, xxxiii. 360.

[2] *Proc. Roy. Soc.*, 1894, p. 132.

[3] Tscherning, *Optique physiologique*, pp. 160 and 168.

[4] *E.g.* cocaine or homatropine.

[5] *Trans. Opt. Soc.*, 1919, xx. 209.

through the front nodal point it does not pass through the centre of the pupil, but through a point about half a millimetre from the centre towards the nose. This is the normal case; in bad cases of excentricity, however, the pupil may be so far off the axis that the ray which represents the line of sight does not enter the pupil at all.

It is difficult to account for the uniqueness of the line of sight and the curious selectiveness by which we can look at *either* of two just separable points and not at *both* simultaneously. It probably arises not from any property of the eye but from the inability of the mind to direct *attention* to more than one point of the image at a time.

The definition falls off very rapidly from the centre of the retina outwards. Few people realise the very bad picture which the eye forms. The fact that we only look at one point at any moment prevents us from realising that the great bulk of the picture is only a suggestive blur; for immediately any part of it attracts our attention we automatically direct the line of sight to it and *fix* it. It is only after some training that an observer is able to give some attention to objects other than the one fixed. This renders very difficult measurements of the properties of non-foveal regions. Only a subordinate degree of attention can be given to the observation, for the primary portion is inexorably demanded by the object at the point of fixation. The moment one gives less attention to this than to the marginal object the line of sight is immediately switched over to the latter, which then ceases to be marginal. This divided attention, combined with the bad definition, renders such observations among the most difficult in physiological optics.

§ (23) The Nature of the Image: Effect of Diffraction.—If optically perfect, and subject only to the limitations imposed by its aperture, the eye could resolve objects 25 seconds apart when the pupil diameter is 5 mm., *i.e.* the image of a point would be a diffraction pattern with a central disc about 50 seconds diameter. According to Hooke, for two luminous points to be perceived separately there must be at least one unaffected cone between those receiving the images. The angular diameter of a cone at the centre of the fovea is about half a minute, so that the images of point sources were really points, the structure of the retina would permit resolution of points about 30 seconds apart. Since, however, the images are discs of appreciable area, the centres of such discs must be further separated in order that there may be a cone between them receiving substantially less light than those on either side.

Hooke found that for very good eyes the minimum separation for resolution is about a minute. This result was confirmed by v. Helmholtz, and is the general experience in the case of acute eyesight.

The size of the diffraction disc is proportional to the wave-length of the light, being nearly twice as great for red as for blue light.

The image of a point is not, however, determined by diffraction alone, being affected by spherical aberration, astigmatism, and chromatic aberration.

§ (24) Spherical Aberration. — The majority of eyes are "under-corrected" for spherical aberration, the marginal portions of the refracting system having shorter focal length than the centre, but various degrees of correction, sometimes even over-correction, are encountered. Frequently there are different degrees of correction in different zones or regions. For instance, aberration may be corrected for one meridian, but under- or over-corrected in the meridian at right angles. Where the pupil is excentrically placed, aberration of opposite signs may be encountered at opposite sides or at the top and bottom. During accommodation for near objects the flattening of the peripheral portions of the lens tends to correct the usual spherical aberration, and may even over-correct it when the degree of accommodation is considerable.

§ (25) Astigmatism.—Nearly all eyes show some astigmatism, or variation of power in different meridians. When at all marked, it results in the image of a point source being no longer a circular disc but a more or less elongated ellipse. Lines parallel and perpendicular to the axes of the ellipse cannot be simultaneously focussed, while lines at intermediate angles cannot be sharply focussed at all. Astigmatism can usually be corrected by means of cylindrical lenses of suitable power suitably orientated with respect to the astigmatic ellipse.

The chief seat of astigmatism is in the cornea, the anterior surface of which may have different curvatures in different directions. The astigmatism is termed *direct* if the greater power is in the vertical meridian, and *inverse* if it is in the horizontal. In other cases it is termed *oblique* astigmatism.

Among a number of persons examined by Nordenson,[1] 9 per cent had no astigmatism, 77 per cent had direct, 1 per cent inverse, and 12 per cent oblique astigmatism. It appears that astigmatism alters with age, the inverse variety becoming more frequent in older people, owing to the increasing tension of the cornea. There is no necessary relation between the astigmatism of the two eyes.

The term *irregular astigmatism* is applied

[1] "Recherches ophtalmométriques sur l'astigmatisme de la cornée," *Ann. d'Oc.*, 1883; *vide* also Tscherning's *Optique physiologique*, p. 117.

to cover any defect of refraction which cannot be corrected by means of a suitable lens. In many cases the image of a point is neither a circular nor elliptic patch, but may be of the most fantastic shapes. If such defects are pronounced the vision is poor and cannot be rendered normally acute with glasses.

We shall see later that astigmatism may seriously affect the judgment in certain optical measurements.

§ (26) Chromatic Aberration.—The eye is only slightly corrected for chromatic aberration. Its chromatic properties are very similar to what they would be if its contents were replaced by water.[1] Usually the actual curves connecting focal length and wave-length are slightly flatter from $\cdot 52\mu$ to $\cdot 66\mu$ than the curves for a simple water eye; but in many cases they are difficult to distinguish. The power for light near the mercury line at $\cdot 436\mu$ is usually about $1\frac{1}{2}$ diopters greater than for yellow light. In ordinary vision the effects of chromatic aberration are not noticed as the eye accommodates so that the brightest part of the spectrum—the yellow green—is in focus, the red and blue rays focussing behind and before the retina. There is therefore a bright image surrounded by a faint purple halo due to the superposed red and blue circles of diffusion. This halo is not noticeable under ordinary conditions; but if only part of the pupil is utilised, as, for instance, when half of it is obscured by a card, effects due to chromatic aberration become manifest. A bright point of light or a luminous line, such as the filament of an electric lamp, appears drawn out into a short spectrum with the red end *apparently* towards the screen, that is with the red rays striking the retina at the other side of the optic axis from the screen. The reason for this is at once apparent when it is borne in mind that the blue rays have crossed the axis before meeting the retina while the red rays have not.

§ (27) Chromatic Parallax.—The experiment just described is even more striking if, instead of employing half the pupil, a slit or "pinhole" about 1 mm. wide is placed in front of the pupil, as far to one side as possible, and two thicknesses of "Cobalt glass" are interposed. The glass transmits a little red light as well as blue, cutting out the central portions of the spectrum. Under these circumstances two separated images of the filament will be seen, one red and one blue. Their positions will obviously be reversed if the hole is moved to the other side of the pupil. Thus on moving the hole to and fro in front of the eye, or moving the eye behind the hole, the red and blue images will move to and fro. This effect has been termed *Internal Chromatic Parallax*,[2] because the separation of the different coloured rays takes place entirely within the eye, the incident pencils being coincident.

If, instead of a single source emitting two kinds of light, we have two sources, one blue and the other red, we may place them at such distances from the eye that both images are focussed on the retina. If the line joining the objects coincides with the line of sight, the images will also coincide. In this case, represented in *Fig.* 7 (*a*), the cones of coloured rays within the eye are coincident. Consequently there is no apparent displacement of the red and blue images on inserting an excentric pinhole in front of the pupil, as in *Fig.* 7 (*b*).

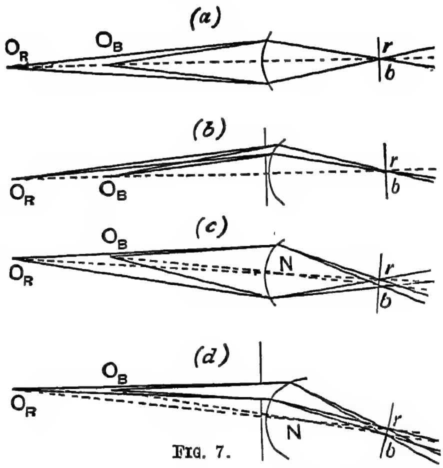


Fig. 7.

If, however, the eye is moved relative to the line joining the objects, a displacement of the images occurs, as is evident from *Fig.* 7 (*c*), each image lying on the line joining its respective object to the nodal point N. Since rays from all parts of the pupil converge to r and b their displacement is unaffected by the interposition of a pinhole; it therefore takes place even if the movement of the eye is behind a fixed stop, *Fig.* 7 (*d*). This phenomenon is termed *External Chromatic Parallax* to distinguish it from the previous case.

It might appear that external chromatic parallax, since it is only evident on moving the eye relative to the line joining two non-coincident objects, is simply the ordinary parallax always observed between objects at unequal distances; but they are in reality quite different, as the effect of a small stop is exactly opposite in the two cases. In the ordinary case when objects at unequal distances from the eye are seen by white light,

[1] P. G. Nutting, *Proc. Roy. Soc.* xc. 440.

[2] Guild, *Proc. Phys. Soc.*, 1917, xxix. 311.

or light of the same colour, the images are also at different distances, the image of the more remote object being in front of that of the nearer object. Also, rays which are coincident in the incident beam are equally refracted, and are therefore coincident within the eye. It follows from these two properties that ordinary parallax will be observed if a small stop is moved across the eye, even if the latter remains fixed, but will not be observed on moving the eye behind a fixed stop. External chromatic parallax, as we have seen, is observed on moving the eye behind a fixed stop, but is not observed on moving the stop before the fixed eye. Like internal chromatic parallax it arises wholly from the different refrangibilities of the two colours.

If we start with two coincident objects of different colours and gradually separate them until they are in the relative positions of *Fig.* 7, the amount of chromatic parallax observable *on moving the stop in front of the fixed eye* will gradually diminish to zero. The amount observable *on moving the eye behind the fixed stop* will be made up of a diminishing amount of internal and an increasing amount of external chromatic parallax. It is easy to show that the total parallax is constant wherever the two objects may be.[1] Chromatic parallax may give serious trouble in spectroscopic work when setting cross-lines on coloured spectrum lines. In well-designed instruments the power of the telescopes employed is about 15 to 20 magnifications for each inch (2·5 cm.) of effective aperture. The "exit pupil" of the telescope is therefore only from 1 to 2 mm. in diameter, and constitutes a small fixed stop through which the objects in the field of view are seen. In making observations on blue and violet lines the observer frequently depends on scattered extraneous light from the brighter portions of the spectrum to provide general illumination for the cross-lines. The cross-lines and spectrum line are therefore seen by light of different wave-lengths, and chromatic parallax effects are produced between them with slight movements of the eye. As we have just seen, this cannot be got over by focussing (that is, altering the relative distances from the eye of cross lines and spectrum line).

Suitable methods of overcoming the difficulty are mentioned in another article.[2]

§ (28) Visual Acuity.—The clearness with which the eye can see detail is termed its acuity. Even when the correctable defects of refraction, such as short or long sight and regular astigmatism, have been corrected by suitable spectacles, there are still considerable variations in the acuity of different eyes.

Acuity is usually tested by means of the Snellen Chart, which consists of a number of rows of letters of different sizes. For unit acuity a row of letters should be legible at such a distance that each letter subtends at the eye an angle of 5 minutes. The lines and spaces composing the letters subtend on an average 1 minute. If a row can just be read when the letters subtend x minutes, the acuity on the Snellen scale is $5/x$. This standard is somewhat low, good eyesight being represented by about 1·5.

The acuity depends on the brightness with which the object is illuminated. Druault found that with a test chart illuminated by 0·016 candle at a metre, the acuity was as low as 0·075. It increased with brightness, rapidly at first and then more slowly. It reached 1·0 for an illumination of 1·5 candle metres, 1·25 at 16·7 and 1·50 at 5400 candle metres. The variation is therefore very slight over an enormous range of brightness.

The diminution of acuity at low intensities is more rapid at the red than at the blue end of the spectrum.

§ (29) Coincidence of Objects in Field of View.—Many important optical and physical measurements depend on the determination of the coincidence of two suitable objects in the field of view of an optical instrument. The setting of telescope cross-lines on spectrum lines, or on the image of a distant staff, are cases in point. With suitable pairs of objects the eye is able to make adjustments of this type with very great precision indeed, provided the necessary precautions are taken to prevent avoidable errors. Many different types of setting are used for various purposes; but it will be sufficient, in order to illustrate the capabilities of the eye in this respect, to quote the accuracy with which want of alignment between two halves of a straight line can be detected. Settings of this kind are met with in the use of scales and verniers, and also in the coincidence type of rangefinder. The field of view of such an instrument is in general divided into two parts by a horizontal line of division. In the upper half one sees the same objects as in the lower half, but upside down. The measurement is carried out by adjusting the two images of the distant object—a flagstaff, for instance—so that one is exactly above the other as indicated by the coincidence of the images at the line of demarcation. J. W. French[3] has investigated the precision with which such settings can be made and the errors to which they are liable. Settings were made with various thicknesses and lengths of line and with various widths of dividing line between the two fields. It was found that under favourable conditions readings could be repeated so closely that the departure of an observation from the mean of a series

[1] Guild, *Proc. Phys. Soc.*, 1917, xxix. 311.
[2] "Spectroscopes and Refractometers."
[3] *Trans. Opt. Soc.*, 1920, xxi. 127.

was only about half a second of arc, or *less than a hundredth part of the displacement necessary for the resolution of two lines.* The precision varies with the angle which the length of the lines subtends at the eye, there being practically no aligning power for lines under 0·6 minute. As the length increases the precision improves up to a length of about 12 minutes, beyond which there is no further improvement. The thinner the lines the more rapidly the final precision is reached, but its actual value is nearly independent of the width of the lines.

The precision is reduced by increasing the width of the horizontal separating line. Thus for a separation of 4 seconds, the limiting value of the mean error was 1·1 seconds, while for a separating line of 19 minutes it was over 5 seconds.

The mechanism by which this close repetition of settings is effected is obscure. In the best case mentioned the two retinal images were evidently always brought into the same relative position to within a sixtieth of the diameter of a cone. This is all the more surprising inasmuch as the setting when made does not really correspond to exact alignment. This is at once revealed by making settings in opposite directions. The difference between such settings is greater the wider the separating line. It may be as little as two seconds for a very fine separation or as much as 40 seconds for a 20 minute separation. Not only so, but the mean of the settings made in opposite directions does not in general give the true setting for actual coincidence. There is thus a "personal equation" which may be many times as great as the maximum variation of individual settings. French traced this error to astigmatism in the eye of the observer. The error is zero when the direction of the astigmatism is parallel or perpendicular to the separating line, and greatest when it is at 45° to it.

It is therefore desirable in making all settings of this nature that the observer should be aware of the direction of any uncorrected astigmatism from which his eye may suffer, and that he should so orient his head as to bring this direction parallel or perpendicular to the lines whose collinearity has to be adjusted.

This aligning power of the eye does not appear to vary greatly with the acuity of vision. Thus French found that in certain experiments the brightness could be reduced 1000-fold without detracting from the precision. The actual illuminations were not stated, but the visual acuity was probably poor at the lower illuminations. This is in accordance with general experience with other types of coincidence setting, in which the definition may become surprisingly poor before the precision of setting is appreciably affected. There is still considerable obscurity as to what property of the eye is responsible for the precision with which these various settings can be made, but it is certain that in many cases the sense of symmetry plays an important part.

J. G.

Eye, Adaptation of the, in Photometry. See "Photometry and Illumination," §§ (31) and (126).

EYEPIECES

§ (1) Introductory.—Ordinary optical instruments intended to be used as aids to vision, other than those designed to correct individual abnormalities in the eye, are for the most part divisible into at least two portions, the first, called the objective, forming a real image of the object to be examined, and the second, which is used in close proximity to the eye, the eyepiece. Eyepieces in normal use always form virtual images of the real image produced by the earlier part of the instrument, though it is to be observed that regarded as an eyepiece object the earlier image may be virtual. As might be expected from the similar conditions desirable in the emergent rays, very similar constructions in the eyepiece are suitable for use in a wide variety of instruments, and from this circumstance the eyepiece has come to be regarded as an independent optical system which can be transferred as a unit from one apparatus to another without detriment to the perfection of the visible image presented to the eye. For many purposes this is sufficiently nearly the case, and the same eyepiece may be used in a laboratory for widely differing work. When extreme conditions are encountered it is no longer the case that the best results are obtainable from an eyepiece of normal type. It is therefore only within limits that eyepieces may be regarded as separate instruments; beyond these limits the most satisfactory eyepieces must have peculiarities in their design which unfit them for general work while making them excellent for the special purpose for which they are intended. In these cases it is essential that the eyepiece should be regarded as an integral portion of the optical system to which it belongs, and the character of the corrections attained by it will depend upon the division of properties between the objective and eyepiece which the designer finds most convenient under the circumstances special to the particular instrument.

§ (2) Huygens' Eyepiece.—The most familiar apparatus in which separate eyepieces are used are the microscope and telescope, and the kinds of eyepiece most frequently

provided with these are known as the Huygenian and the Ramsden. In their simplest forms both are built of two separated plano-convex lenses. In the Huygenian eyepiece (*Fig.* 1) both lenses present their convex faces to the incident light, their separation is very nearly equal to half the sum of the focal lengths of the components, and the lens nearer the objective, called the field lens to distinguish it from the second lens, which is called the eye lens, is always of greater focal length than the eye lens. For eyepieces of short focal length, suitable for procuring high magnifications, the ratio of the two focal lengths is about 3 : 1, and as the focal length increases this ratio diminishes, until for the

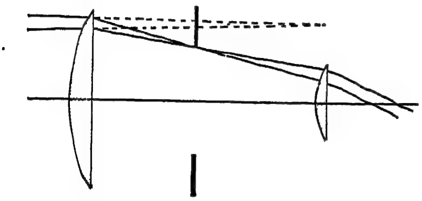


FIG. 1.—Huygens' Eyepiece.

eyepieces of greatest focal length it may be about 3 : 2. Since this inequality involves a real focus for incident parallel light travelling in the reverse direction through the instrument before the field lens is reached, and consequently a virtual final image for that light, this eyepiece cannot be used for the direct examination of a real object, and by analogy with simple lenses has therefore been termed a negative eyepiece. The term is quite inappropriate and misleading, for in so far as "positive" and "negative" are at all suitable for application to optical systems in general they must relate to the sign of the power, which is in this case positive as for each component lens. In fact, if the ratio of the focal lengths is $m : 1$, the elements from which a system of power κ must be constructed are

$$\text{Power of field lens } \frac{2\kappa}{m+1},$$

$$\text{Power of eye lens } \frac{2m\kappa}{m+1},$$

$$\text{Separation } \frac{(m+1)^2}{4m\kappa}.$$

The positions of the principal foci are distant respectively from the field and eye lenses

$$-\frac{m-1}{2\kappa} \text{ and } \frac{m-1}{2m\kappa},$$

the negative sign in the former case indicating that the focus is virtual.

§ (3) RAMSDEN'S EYEPIECE.—In Ramsden's eyepiece (*Fig.* 2) the two lenses are of equal focal length, their separation being somewhat less than the focal length of either, generally from $\frac{3}{4}$ to $\frac{2}{3}$. The most obvious difference,

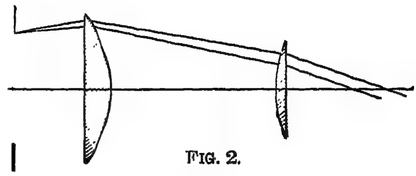


FIG. 2.

however, is that the field lens is reversed, thus presenting its plane side to the objective. If the ratio of the separation to the focal length of a component is n, the data for the construction of a system of power κ are

$$\text{Power of either component} = \frac{\kappa}{2-n},$$

$$\text{Separation} = \frac{n(2-n)}{\kappa}.$$

The principal foci lie outside the lenses at clear distances of approximately $(1-n)/\kappa$ and the eyepiece is very convenient for use in examining a real object. In particular, this property causes it to be generally used in conjunction with a thread micrometer for fine measurements of position.

§ (4) EYEPIECE THEORY.—These two eyepieces seem to have been evolved without the guidance of any adequate theory, and it is thus of interest to consider how they compare with the forms to which present theories would lead. It is to be remarked in the first place that for the attainment of a large field of view the presence of a field lens is necessary, the effect being to bend towards the eye lens rays that would otherwise reach the transverse plane in which this lens is situated at too great a distance from the axis to be transmitted out of the instrument. The eyepiece must thus consist of at least two well-separated lenses, and these will evidently both be of positive power. Another important consideration is that the eye is external to the instrument, and attention is directed to various parts of the field of view by rotating the eyeball in its socket without relative movement of the instrument and the observer's head. The useful rays, therefore, lie within cylinders which pass through the rim of the pupil and have axes through the centre of rotation of the eyeball. It follows that the useful rays from different object-points traverse different parts of the eyepiece, and each separate beam is about the diameter of the pupil, and occupies only a small part of the lens aperture. Under these conditions defects such as spherical aberration and coma, which are of outstanding importance for objectives, become of little moment in

comparison with the defects which do not depend on the width of the beam, such as curvature and astigmatism. Accordingly in eyepieces used under normal conditions, as for high-power telescopes, the primary requisite is a good balance in the curvature corrections, the eyepiece being required to correct the errors of this kind of the objective in addition to its own tendency to produce such defects, while the objective must if necessary be left with small outstanding errors of spherical aberration and coma to correct those of the eyepiece. As is well known, the corrected objective causes the rays to form the focal lines of the issuing beams on two surfaces of curvatures $-\kappa'(\delta'+\varpi')$ and $-\kappa'(3\delta'+\varpi')$, where κ' is the power of the objective and ϖ' is its Petzval coefficient, the value of which is about 0·7.

δ is the quantity denoted by δ_3 in § (17) of the article on "Optical Calculations," and on which the curvature of the image depends. For a lens free from spherical aberration and coma it may be shown that δ is unity. The quantity $\kappa\varpi$, which is given by the equation

$$\kappa\varpi=\Sigma\left(\frac{1}{\mu_{p-1}}-\frac{1}{\mu_p}\right)\frac{1}{r_p},$$

is known as the Petzval sum.[1] κ is the power of the system, and ϖ the Petzval coefficient, which is clearly a pure number. The above expressions should be compared with the quantities $-(\epsilon+\epsilon')/v_3$ and $-(3\epsilon+\epsilon')/v_3$ of the article on "Lens Systems, Aberrations of," and "Photographic Lenses."

Thus for a corrected objective the curvatures are $-\kappa'(1+\varpi')$ and $-\kappa'(3+\varpi')$. These errors would be entirely overcome if the corresponding curvatures of the eyepiece, $-\kappa(\delta+\varpi)$, $-\kappa(3\delta+\varpi)$, satisfied

$$\kappa'(1+\varpi')+\kappa(\delta+\varpi)=\kappa'(3+\varpi')+\kappa(3\delta+\varpi)=0,$$

or $$\kappa'+\kappa\delta=\kappa'\varpi'+\kappa\varpi=0.$$

In fact, the second of these conditions cannot be met, for κ, κ', ϖ, ϖ' are all necessarily positive. The best that can be done is to effect a compromise, and there is no general agreement on the precise form this should take. If the condition

$$\kappa'+\kappa\delta=0$$

is taken, the complete instrument will be corrected for astigmatism, but the field of view will appear to be convex to the observer. As κ' is small in comparison with κ in the cases we are considering, this condition implies a small negative value for δ. The outstanding curvature can be removed at the expense of astigmatism by making

$$\kappa'(2+\varpi')+\kappa(2\delta+\varpi)=0,$$

δ then having a larger negative value. Any particular type of correction may be obtained by securing an appropriate value for δ, but the values corresponding to all types of correction aimed at in eyepieces are invariably negative.

As the contributions to the curvature terms of the various lenses in an instrument are directly additive, it is seen that the condition is most readily met by securing negative contributions from both the field lens and the eye lens. From the result quoted in the case of the objective it appears that the presence of spherical aberration or coma in the eyepiece is a condition essential to the attainment of the required curvature correction.[2] Treating both lenses in the eyepiece as thin, the value of δ, apart from a coefficient which is necessarily positive, is

$$\gamma-2\beta(\mathfrak{S}+S)+2(1+\varpi)\mathfrak{S}S+S^2,$$

or if $s\kappa$ is the curvature added to each surface in deriving the shape of the lens from its standard form,

$$\gamma_0-2\beta_0(\mathfrak{S}+S)-\frac{1}{1+2\varpi}\{(1+\varpi)\mathfrak{S}-\varpi S\}^2$$
$$+\frac{1}{1+2\varpi}\{2(1+2\varpi)s-(1+\varpi)(\mathfrak{S}+S)\}^2.$$

Now for a single thin lens of refractive index μ

$$\varpi=\frac{1}{\mu},\quad \beta_0=0,\quad \gamma_0=\frac{1}{(1-\varpi)^2},$$

and with given values of $\mathfrak{S}$ and S this expression is algebraically a minimum when

$$2(1+2\varpi)s=(1+\varpi)(\mathfrak{S}+S).$$

Consider in the first place the eye lens. For simplicity the direction in which the light travels may be taken as the reverse of that in a complete instrument, so that we are dealing with parallel incident light and a front stop not far from the lens. These conditions give $\mathfrak{S}=1$, while S is negative and comparatively large. The minimum value is thus attained for a negative value for s, so that the surface of greater curvature is towards the objective. The limiting case when the contribution of the eye lens to δ is zero evidently occurs in a single lens when

$$s=\frac{1+\varpi}{2\varpi}\left\{\mathfrak{S}\pm\left(\frac{\gamma_0}{1+2\varpi}\right)^{\frac{1}{2}}\right\},$$

and the shallowest curvatures will be obtained by choosing the negative sign before the square root. If the glass is of refractive index 1·5, so that $\varpi=\frac{2}{3}$, the preferable solution is approximately

$$s=-1\cdot2.$$

[1] This may also be written as $\Sigma\varpi_m\kappa_{m,m+1}$. See "Optical Calculations," § (7), equation (39).

[2] For the values of the quantities β, γ, δ, see "Optical Calculations," § (7), equations (42), (45); § (8), equations (48), (49); and § (17); δ corresponds to the curvature coefficient δ_3 of that article. See also above.

so that the eye lens is meniscus with radii in the ratio 11 : 1, the corresponding value of $\mathfrak{S}$ being $-11/4$. For manufacturing purposes it is preferred to make the outer surface of the eye lens plane, and the small change thus introduced is easily compensated by a slight alteration in the value of $\mathfrak{S}$.

The discussion of the desirable shape for the field lens may follow similar lines. In this case $\mathfrak{S}$ is small and may usually be entirely neglected. The best conditions then require s and $\mathfrak{S}$ to agree in sign, that is to say, the greater curvature in the field lens must be on the side away from the real image. This simple theory thus accounts satisfactorily for the shapes of the lens when simple glasses are used. It will be noted that the determining factor is the positive value of γ_0. Since for ordinary achromatic lenses γ_0 is negative, it would be expected that the substitution of such lenses for the simpler lenses would be unsatisfactory, and this accounts for the unsatisfactory performance of the Kellner

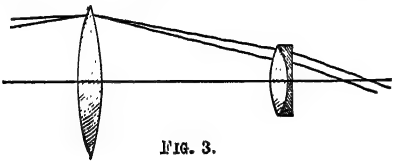


FIG. 3.

eyepiece (*Fig.* 3), which is a modified Ramsden with achromatic eye lens and double convex field (or sometimes plano-convex) lens situated in the plane of the real image. The unsatisfactory features of this objective could be removed by making use of a dense barium crown glass instead of a hard crown in the cemented eye lens, the high refractive index of the crown glass resulting in a suitable positive value for γ_0.

The method of correction which has been outlined obviously provides for the removal of spherical aberration and coma, and a large reduction in the amount of curvature and astigmatism. Of the spherical aberrations there remains distortion. The eccentric paths of the principal rays through the eyepiece inevitably result in the presence of distortion in the visible image, so that straight lines in the object are not represented by straight lines in the image. As long as the defect is not very obvious it is of no consequence, and when a Ramsden eyepiece is employed even pronounced apparent distortion is of no importance for exact measurements. This is evident when it is recollected that the image produced by the objective is free from distortion and that it is actually this image which is measured, the apparent distortion affecting the image and the measuring device equally. It is, however, of interest to remark that if the mathematical conditions for the removal of distortion were satisfied the image would appear distorted. This is because the judgment of the eye in examining the image is influenced by the presence of the unwonted circular boundary by which the visible field is limited, in consequence of which the apparent absence of distortion corresponds in fact to the presence of a very appreciable amount of real distortion.

§ (5) CHROMATIC DEFECTS.—In addition to the removal of the errors which have been mentioned, eyepieces must present to the eye images which appear to be corrected for colour. The complete correction of the system for colour would involve the employment of achromatic field and eye lenses, in addition to an achromatic objective. The insensitiveness of the eye under ordinary conditions to limited errors renders such elaboration unnecessary, and satisfactory results may be obtained for limited fields with the single glass lenses, the cost of which is naturally very much less than that of achromatic lenses. The use of the simple constructions means that only one colour condition can be satisfied. When the final image is at infinity, obviously it is only necessary that the images of different colours should subtend the same apparent angles at the eye. More generally the desirable condition is that the principal ray for a secondary colour should pass through the image for the fundamental colour. Consider now the properties of the Huygenian eyepiece for light of a colour for which the refraction is increased by one part in ν above that for the colour previously considered. The power of each lens is then obtained by multiplying the old value by $1+1/\nu$. The result is readily seen to be that the focal length of the combination is κ as for the original colour, but the focal planes are moved, their distances from the external surfaces being now

$$-\frac{\{m-1+(m+1)/\nu\}}{2\kappa}$$

and

$$\frac{\{m-1-(m+1)/\nu\}}{2m\kappa}.$$

The suggested condition for achromatism is thus met if the objective is at a great distance from the eyepiece when the emergent light is parallel. This is the condition appropriate in high-power telescopes, but for instruments in which the objective is not very distant the best results are obtained by a modified construction, and the altered eyepiece is known as a compensating eyepiece.

In the case of the Ramsden eyepiece the power is not the same for the new colour, the new value being

$$\kappa\left\{1+\frac{2(1-n)}{(2-n)\nu}\right\},$$

and the distance of the principal focus from the nearest lens is

$$\left\{1-n-\left(\frac{2}{2-n}-n\right)\frac{1}{\nu}\right\}\frac{1}{\kappa},$$

and the signs in the two changes are such that the desirable colour correction cannot be effected. The compensation required in the case of the Ramsden eyepiece is thus of a heavier character than with the Huygens, and this may be the reason why of the simpler eyepieces the Huygens is so decidedly favoured where it is possible to employ it, as, for example, in microscopy. For precise work where the Ramsden form is employed one lens is usually compound, an arrangement which enables the desirable conditions to be achieved with ease.

§ (6) Microscope Eyepieces.—The modifications which have been introduced into these lenses either consist in the substitution of a field lens of flint glass for the crown, or in a more elaborate construction. One modification of the Ramsden type has already been described. Another is the form used by Zeiss (*Fig.* 4), illustrated in the accompanying figure.

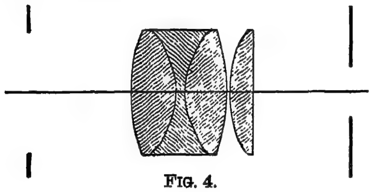


Fig. 4.

For microscopy it is frequently desired in high-power work to use a low-power eyepiece in the first place, and substitute another of high power for more critical examination of the specimen. This interchange is much facilitated if refocussing can be avoided, and Zeiss introduced the modern system of microscope eyepieces in which the image plane is in a corresponding position for all members of the series as they are placed in the draw tube. This convention is now generally followed.

When the magnification is very high the field of view is correspondingly small, and the need for a field lens disappears. Accordingly a number of high-power eyepieces are made of simple achromatic cemented lenses. For these very small fields the correction for curvature remains important, because the very powerful apochromatic objectives which are invariably used for such work suffer from curvature to a much greater extent than do objectives of other types.

§ (7) Erecting Eyepieces.—The eyepieces which have so far been described all yield an inverted image of an external object. For a telescope used for terrestrial objects this is unacceptable, and the image must either be erected by the use of reflecting prisms, in which case a Ramsden eyepiece is suitable, or an erecting eyepiece must be used. The latter consists, as a rule, of four separated lenses, and in the simplest cases these are all lenses of a single glass. When large fields of view are involved at least one of the lenses is of more complex construction. The first erecting eyepiece is due to Dollond, and the best form has been investigated by Sir George Airy (*Fig.* 5). The eyepiece which he recommends is illustrated in the figure, and the data he gives are as follows:

Lens.	Focal Length.	Shape.	Ratio of Radii.	Separations.
1	3	double convex	6 : 1	4
2	4	meniscus	25 : 11	6
3	4	convexo-plane	. .	5·13
4	3	double-convex	1 : 6	

The third and fourth lenses constitute a modified Huygens eyepiece, and their distance from the first two lenses is modified to remove any observable trace of colour in the image.

Airy also investigated in detail suitable forms for the Huygens and Ramsden eyepieces. For the former he concluded that the most desirable construction is

Field lens, $f=3$, meniscus; ratio of radii, 4 : 11,
Eye lens, $f=1$, double convex; ratio of radii, 1 : 6,

with a separation of 2.

§ (8) Special Forms of Eyepiece.—Many variations of these forms of eyepiece have been described, and there is no doubt that excellent results are obtainable by a great variety of constructions, provided these are more complicated than those which have been described here. There is, however, little to be gained by a discussion of such forms, for the best results in all cases will be obtained by considering the complete instrument as a whole. An idea of the forms assumed in cases where this is done may be obtained by reference to the article on "Telescopes." The most notable feature of the newer systems is that every effort is made to avoid the use of lenses for erecting the image, the employment of prisms for this purpose being highly preferable because the image is thereby more easily susceptible of correction for curvature and astigmatism. Whenever possible there is much to be said for the use of simple eyepieces, the brilliancy of the final image varying

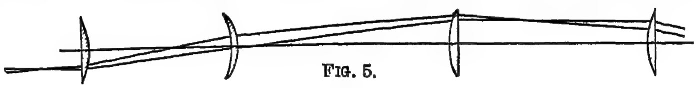


Fig. 5.

inversely with the number of lenses used in the system.

It will be evident from the trend of the foregoing discussion that the circumstances which call for special modifications of eyepiece are abnormally large fields of view and particularly low magnifying powers. In the latter case the conditions for the correction of the image are so different from those at high powers that it may easily occur that the least expensive construction will result from the use of an eyepiece over-corrected for colour and an uncorrected objective rather than the usual reverse arrangement.

The so-called diagonal eyepiece is obtained by inserting an inclined mirror into the path of rays in the eyepiece, the emergent direction of the light being thus inclined, usually at right angles, to the incident direction. It is employed only where a direct eyepiece is inconvenient.

For polarising eyepieces and autocollimating eyepieces see articles on "Microscope" and "Spectroscopes and Refractometers."

§ (9) Variable Power Eyepieces.—An entirely distinct class of eyepiece from any of the foregoing comprises the variable power systems. Their motion is controlled by the rotation of a sleeve in which suitably shaped slots are cut, these serving to give longitudinal movements to the lenses through the medium of feathers which engage in the slots. The lens carriers are, of course, prevented from rotating about the axis of the instrument. In at least one form there is no longitudinal movement of the eye lens as the power is changed, so that the over-all length of the instrument is constant and it becomes more simple than in other cases to prevent the ingress of water or other objectionable material into the interior of the instrument.

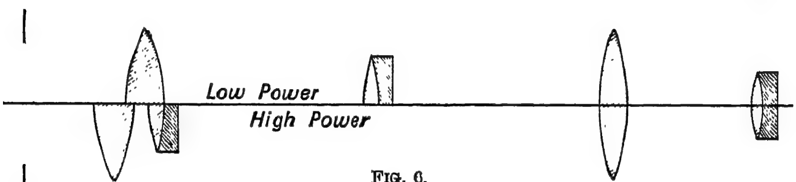


Fig. 6.

The moving systems of lenses which are used in some eyepieces of this class are more allied to photographic lenses than to most other optical systems, and the parallel in the work that has to be performed in the two cases will be realised. The diagrams show two eyepieces of this class, one made by Messrs. Ottway (*Fig.* 6), and the other by Messrs. Ross (*Fig.* 7).

The possibility of constructing a system with two moving parts to give a constant distance between object and image at various magnifications may be seen by considering a system of two lenses of focal lengths f_1 and f_2,

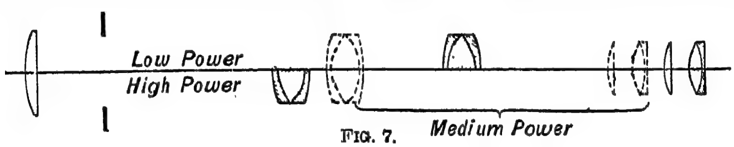


Fig. 7.

eyepieces, in which, by a suitable movement, the extent to which the image is magnified can be varied, the view remaining in good focus throughout the change. The ratio of the highest to the lowest magnifying power is usually from 3 : 1 to 4 : 1. The conditions which have to be fulfilled naturally vary considerably from one another at the extreme ends of the range, the correction of curvature being of outstanding importance at one end, while at the other the extent to which spherical aberration and coma can be tolerated is more limited. The lenses employed necessarily include achromatic combinations, and the power is varied by introducing large changes in the separations of the components. To avoid large changes in the external length of the instrument while maintaining constant focus there must be at least two independent moving the separation of whose inner principal surfaces is t, where

$$t = L \pm \{(L - f_1 - f_2)^2 - (f_1 + f_2G)(f_1 + f_2/G)\}^{\frac{1}{2}}.$$

When the magnification is G, the distance d_1 from the first principal point of the first lens to the object is

$$d_1 = \frac{f_1}{f_1 + f_2G}[L - f_2(1 - G) \mp \{(L - f_1 - f_2)^2 - (f_1 + f_2G)(f_1 + f_2/G)\}^{\frac{1}{2}}],$$

and similarly, the distance d_2 from the second principal point of the second lens to the image is given by

$$d_2 = \frac{f_2}{f_1 + f_2G}[LG + f_1(1 - G) \mp G\{(L - f_1 - f_2)^2 - (f_1 + f_2G)(f_1 + f_2/G)\}^{\frac{1}{2}}],$$

so that $$d_1 + t + d_2 = 2L.$$

These equations suffice for the theory of position of the moving parts when the over-all length is not restricted to a constant value if L is regarded as a variable quantity. It will be observed that in any case the magnification is determined by the focal lengths f_1 and f_2, and the distances of the object and image irrespective of the values of t and L, for

$$\frac{d_2}{f_2}-1=G\left(\frac{d_1}{f_1}-1\right).$$

As the moving lenses form the erecting portion of the eyepiece, G will be negative, so that for any magnification one lens produces a virtual and the other a real image. In the special case $d_1=f_1$, $d_2=f_2$, G is given by the ratio of the two focal lengths. T. S.

EYEPIECES, CORRECTIONS FOR OPTICAL DEFECTS. See "Eyepieces," § (4).
Chromatic Defects. See *ibid.* § (5).

EYEPIECES FOR TELESCOPES. See "Telescope," § (6).

— F —

FATIGUE, VISUAL. See "Eye," § (16).

FECHNER'S LAW: a law governing the relation between the intensity of light and the degree of sensation produced. See "Eye," § (5).

FELSPAR ($K_2O_1Al_2O_36SiO_2$), USE OF, IN GLASS MANUFACTURE. See "Glass," § (5) (iii.).

FIGURE OF THE EARTH. See "Surveying and Surveying Instruments," § (6). See also "Gravity Survey," Vol. III.

FILMING: formation of a surface layer interfering with transparency on the interior surfaces of lenses and prisms in optical instruments. See "Glass, Chemical Decomposition of," § (1).

FILMLESS PHOTOGRAPHY. See "Graticules."

FILMS, THIN, INTERFERENCE OF LIGHT IN. See "Light, Interference of," § (7).

FILTER (LIGHT): a transparent slab of coloured material, either liquid or solid, used to impart colour to the light which it transmits. See "Eye," § (13).

FILTERS:
Compensating light: used in connection with the screen plate processes of colour photography. See "Light Filters," § (2) (iv.).
Contrast light, description and use of, in photography. See *ibid.* § (2) (ii.).
Orthochromatic light, description and use of, in photography. See *ibid.* § (2) (i.).
Selective light, description and use of, in colour photography. See *ibid.* § (2) (iii.).

FINDER, PHOTOGRAPHIC: a device used for the purpose of indicating the amount of subject which would be included on the focussing screen of a given camera placed in the same position. See "Photographic Apparatus," § (7).

FIXATION, VISUAL: the act of looking directly at an object in the field of view. See "Eye," § (22).

FLAME, SENSITIVE: a coal-gas flame, burning under certain conditions of pressure and sensitive to sounds of high or medium pitch. See "Sound," § (55).

FLARE SPOTS IN CAMERA LENSES. See "Camera Lenses, Testing of," § (5).

FLARES, PHOTOMETRY OF. See "Photometry and Illumination," § (122).

FLICKER, THE ELIMINATION OF: a problem of great importance in the intermittent type of kinematograph, dependent upon the duration of the individual impulses that fall upon the retina. See "Kinematograph," § (3).

FLICKER PHOTOMETER: a device for comparing lights of different colours. See "Photometry and Illumination," § (95) *et seq.*

FLUTE: a wood-wind musical instrument. See "Sound," § (34).

FLUX (LUMINOUS): the rate of flow of radiant energy evaluated according to its capacity to produce the sensation of vision. See "Photometry and Illumination," §§ (1) and (2).

FOCAL LENGTHS, DETERMINATION OF. See "Objectives, Testing of Compound."
Autocollimating method. See *ibid.* § (2) (i.).
By conjugate points. See *ibid.* § (2) (ii.).
Magnification methods. See *ibid.* § (2) (iii.).
Microscope methods. See *ibid.* § (2) (iv.).
Using parallel light. See *ibid.* § (2) (i.).
See also "Camera Lenses, Testing of," § (1).

FOCAL SPHERE: a term used in the study of projection apparatus to denote the smallest possible sphere described round the "focal point" so that it just, and only just, includes all the convergent rays, either intersecting it or just touching it. See "Projection Apparatus," § (4).

FOCOMETER: a device for enabling the crater of an arc to be maintained in position at the focus of a projector mirror. See "Photometry and Illumination," § (121).

FOG-HORN: an arrangement for producing a loud noise, erected on a cape or other important point, for the warning and guidance of mariners during fogs. See "Sound," § (51).

FOOT-CANDLE: the unit of illumination on the F.P.S. system. See "Photometry and Illumination," § (2).

FOOT-CANDLE METER: a portable illumination gauge. See "Photometry and Illumination," § (63).

FORBES RANGEFINDER. See "Rangefinder, Short-base," § (7).

FORCED LIFE TEST (OF ELECTRIC LAMPS). See "Photometry and Illumination," § (80).

FORK, TUNING-: a convenient standard of musical pitch, for ready reference, each prong being like a *fixed-free* bar and emitting practically its prime partial only, when bowed lightly and carefully at or near the ends of the prongs. See "Sound," § (49).

FORKS, TUNING-, COMPARED BY SMOKE TRACES. See "Sound," § (53) (iv.).

FORREST ARC: a particular form of carbon arc proposed as a primary standard of light. See "Photometry and Illumination," § (11).

FOUCAULT'S HELIOSTAT: a clockwork device for directing the rays from the sun into a fixed telescope or other optical system. See "Telescope," § (17).

FOUCAULT'S METHOD OF DETERMINING POSITION OF AN IMAGE PLANE. See "Objectives, Testing of Compound," § (1).

FOUCAULT'S SHADOW METHOD OF DETERMINING THE CURVATURE AND QUALITY OF CONCAVE SURFACES. See "Spherometry," § (10).

FOUCAULT'S TEST FOR THE MIRROR OF A REFLECTING TELESCOPE. See "Telescope," § (11).

FRAUNHOFER LINES: the name given to the dark lines crossing the solar spectrum, which were first mapped by Fraunhofer. See "Wave-lengths, The Measurement of," § (1).

FREQUENCIES, COMPARISON OF, by monochord method. See "Sound," § (53) (vii.).

By resonance tube. See *ibid.* § (53) (vi.).

FREQUENCIES, EXPERIMENTAL DETERMINATION OF DIFFERENCE BETWEEN, by beats. See "Sound," § (53) (i.).

FREQUENCIES, RATIO OF, gauged by the ear. See "Sound," § (53) (iii.).

FREQUENCY, EXPERIMENTAL DETERMINATION OF. See "Sound," § (53).

FREQUENCY OF A MUSICAL NOTE: a term used to denote the number of vibrations per second occurring in the note. See "Sound," § (1).

FRINGES, ACHROMATIC: produced by using as the two sources of light a short spectrum and its virtual image formed by reflection in a glass plate. See "Light, Interference of," § (5).

FRINGES, INTERFERENCE, METHODS OF PRODUCING, depending on the formation, by some optical system, of two real or virtual images of a narrow source of light, which images act as secondary sources. See "Light, Interference of," § (4).

FURNACES FOR MELTING GLASS. See "Glass," § (15).

— G —

GASES, SPECIFIC HEATS OF, APPLICATION OF QUANTUM THEORY TO. See "Quantum Theory," § (6).

GAS-TUBE: an X-ray tube which depends for its action on the presence of the residual gas in the tube. See "Radiology," § (12).

GEOMETRICAL THEORY OF TELESCOPES. See "Telescope," § (1).

GHOSTS: a name given to spurious lines which occur in all grating spectra and are due to periodic errors in the ruling. See "Wave-lengths, The Measurement of," § (2).

GLARE (in Illumination). See "Photometry and Illumination," § (71).

GLASS

I. DEFINITION AND CONSTITUTION

§ (1).—No satisfactory, concise definition has been suggested for the term glass, which we use, commonly, to denote vitreous, apparently amorphous, materials, brittle at ordinary temperatures, having a high softening point and possessing the general characteristics of an under-cooled liquid.

There is, however, a gradual accumulation of evidence which seems to suggest that glass, under normal conditions, consists, not merely of a single liquid phase, but of one or more solid phases in fine suspension in the liquid phase or, possibly, in colloidal solution.

We may regard glass, at high temperatures, as being a mutual solution of silicates or of oxides. As the temperature falls we approach a point where the glass becomes saturated with respect to one or more compounds stable at that temperature, and we might expect crystallisation to begin. The viscosity of glass, however, increases very rapidly with a fall in temperature (see "Annealing," § (19)), and it generally happens that the viscosity is too high and the rate of diffusion too low to admit of extensive crystal growth during the short time occupied by the cooling of the glass. If glass is cooled excessively slowly the crystals become visible to the eye: if cooled rapidly no visible crystals are formed, but it is improbable that the viscosity will totally inhibit the formation of crystal nuclei and incipient crystallisation when the saturation point of the glass with respect to one of its constituents is reached. (See also under "Physical Properties," § (21).)

Some glasses, which can be worked in the blow-pipe without crystallisation if they have never been allowed to cool (*i.e.* if the glass is gathered straight from the furnace), crystallise at once if worked after having been cooled and re-heated in the blow-pipe. This suggests that, during the cooling, crystal nuclei have been formed, and these tend to grow as soon as the viscosity falls.

It is found that glasses on cooling show a small evolution of heat when passing through a particular temperature. (This is reversible, an endothermic reaction occurring on heating.) The evolution and absorption is extremely small, and can only be detected with the most sensitive apparatus, but is sufficiently established to point to a small but definite change in the properties of the glass at this temperature.

Interesting evidence bearing on the crystalline nature of glass has been obtained by a study of the phosphorescence of glass when exposed to rays of short wave-length. Many glasses show phosphorescence to a marked extent when exposed to ultra-violet light or to cathodic discharge, and it is found that devitrified glass is strongly phosphorescent. If such a glass, however, is cooled rapidly from a high temperature the phosphorescence decreases, and, if chilled sufficiently to form the well-known phenomenon of a Rupert's drop, the skin appears to be free from phosphorescence although the centre of the drop still glows faintly. It seems probable, therefore, that glass in general contains crystalline nuclei in various stages of development, depending on the composition of the glass and the treatment it has received during manufacture. The effect of the crystalline phase will be considered below under different headings.

II. Historical Note

§ (2).—The discovery of glass has been ascribed both to Syria and to Egypt, and there is still some doubt as to the locality and date of the first glass-makers.

It is certain, however, that the Egyptians in 1400 B.C. were acquainted with the art of glass-making, and specimens of glass, blown with and without moulds, have been found dating from that time. They were also familiar with the technique of cutting and the production of coloured glasses. The art of glazing pottery-ware dates back probably to 4000 B.C.

The Greeks and Romans doubtless learned the art of glass-making from the Egyptians, but the glass industry never appeared to thrive in Greece, probably because of the excellence of their ceramic ware. The Romans, however, acquired considerable skill and were successful in mastering the technology of coloured glass, both clear and opal, although they continued for some time to import glass-ware from Sidon, Tyre, and Alexandria. The Romans worked with a glass of the soda-lime type. The Venetians may be credited with being the first to produce glass at a cost which allowed of its more general use. Venetian glass was characterised by the great beauty of form and lightness of the ware. Since the sixteenth century the Bohemians have also been great glass-makers.

The glass industry was first introduced into England in the thirteenth century, when French glass-workers settled in Surrey and Sussex. The industry flourished until the reign of Elizabeth, when restrictions on the use of timber as a fuel caused its temporary decline. In 1550 a factory making window-glass and drinking-vessels was started in London, and in 1619 the glass industry of Stourbridge sprang up. Stourbridge was particularly suited for glass-making owing to the presence of coal, together with the necessary clay for the manufacture of pots. The glass industry has now spread to all parts of England, but Stourbridge still remains one of the chief centres of the flint-glass industry.[1]

III. Composition

§ (3).—The following are the most common glass-forming oxides:

Acidic: Silica; Boric acid.

Basic: Sodium oxide (Na_2O); Potassium oxide (K_2O); Oxides of Barium, Calcium, Magnesium, Zinc, Manganese, Lead, Aluminium, Iron.

In addition to these, the oxides of arsenic, lithium, tin, and zirconium, and the elements fluorine and selenium, are introduced into

[1] See *Encyclopædia Britannica*, 11th edition.

glass for special purposes, and, for the manufacture of coloured glasses, the oxides of nickel, cobalt, chromium, copper, uranium, and the elements carbon, sulphur, silver, and gold are employed.

Probably over 99 per cent of the glass manufactured in the world is composed of a combination of silica with a mixture of two bases, an alkali (soda or potash), and either lead oxide or lime, 90 per cent being made from silica, soda, and lime only. Window glass, plate glass, and bottle glass-ware are made from silica, soda, and lime. Alumina is occasionally added to give greater strength and stability. Bohemian glass-ware is made from silica, potash, and lime, and flint glass for table-ware, cut glass, electric light bulbs, etc., contains silica, lead, and soda or potash. Occasionally magnesia or baryta is substituted for lime or baryta for lead oxide.

For glasses requiring special properties, resistance to attack by chemical reagents,[1] resistance to sudden changes of temperature, etc., magnesia, zinc oxide, and alumina are introduced and boric acid (B_2O_3) is substituted for part of the silica. Glasses free from silica are found to be insufficiently stable for general use. In following column is given a Table of Analyses of typical glasses.

	SiO_2.	B_2O_3.	Na_2O.	K_2O.	CaO.	MgO.	BaO.	ZnO.	Al_2O_3.	Fe_2O_3.	PbO.	As_2O_3.	P_2O_5.	MnO.	Sb_2O_3.	Total.
Resistance glass: Jena	64·60	8·7	9·71	tr.	tr.	0·32	..	10·43	6·24	tr.	..	..	..	tr.	..	100·00
Resistance glass: English	66·51	4·57	11·52	2·58	4·35	0·33	..	3·62	6·74	0·08	..	..	..	0·1	..	100·40
Resistance glass: Köln-Ehrenfeld	66·80	4·13	7·40	1·75	1·73	2·60	..	9·75	2·54	0·17	..	2·05	..	tr.	1·29	100·21
Resistance glass: "Pyrex"	84·62	11·9	3·83	0·61	0·22	0·29	..	..	2·00	0·16	..	0·66	..	tr.	..	100·29
Thermometer glass: Jena 59‴	72·86	10·43	9·82	0·10	0·35	0·20	..	..	6·24	tr.	..	..	..	tr.	..	100·00
Thermometer glass: Jena 16‴	66·58	0·91	14·80	tr.	7·18	0·17	..	6·24	3·84	..	..	..	..	0·28	..	100·00
English lead flint	53·18	..	0·42	11·48	0·14	0·32	..	..	0·50	tr.	33·74	..	..	tr.	..	99·78
Plate glass	71·56	..	12·01	tr.	15·18	0·28	..	..	0·44	0·26	..	..	..	0·20	..	99·93
Tubing for lamp working 1	69·30	..	12·80	4·96	8·28	0·07	..	..	4·05	0·18	..	tr.	..	0·20	..	99·84
Tubing for lamp working 2	69·64	..	13·36	6·98	6·40	0·37	..	..	3·02	0·06	..	..	..	0·18	..	100·01
Common bottles 1	73·0	..	12·09	..	11·5	1·81	..	..	0·90	0·44	..	..	..	0·30	..	100·04
Common bottles 2	75·85	..	15·52	..	7·68	0·12	..	..	0·77	0·05	..	..	..	..	..	99·99
Common bottles 3	65·70	..	8·18	..	16·06	1·08	..	..	2·89	3·67	..	..	..	2·30	..	99·88
Combustion tubing 1	66·9	7·22	1·25	2·40	7·94	0·61	7·27	..	6·38	0·22	..	..	..	..	..	100·19
Combustion tubing 2	79·57	..	0·66	11·60	7·80	0·11	..	..	0·32	0·04	..	..	..	..	..	100·1
Jena lamp glass	73·88	16·48	6·67	..	..	..	..	..	2·24	..	..	0·73	..	..	..	100·00

IV. Raw Materials

The sources of raw materials available for glass-making depend largely on the type of glass for which it is intended, since the small amount of impurities which can be permitted in better qualities of glass excludes a large quantity of material which would otherwise be suitable.

§ (4) Silica is almost invariably introduced into glass in the form of sand. Deposits of sand suitable for glass-making are found in many parts of the world. The chief requirements for a glass-making sand are freedom from iron and evenness and angularity of grain. For the manufacture of bottles, etc., uniformity in grading is the most important feature, the iron content being of less importance, whereas for "white" glass (or "flint" glass, the name now used to denote glass free from colour), such as chimneys, pressed ware, table glass, etc., chemical purity is equally essential. 90 per cent of the sand grains should lie between 0·5 and 0·1 mm. in diameter. Many sands are found containing 99 per cent between these limits. Sands graded between smaller limits 0·5 and 0·25 mm. are to be preferred, and it is not uncommon to find sands containing over 50 per cent between these limits. In America graded sands considerably coarser than this are sometimes used.

[1] See article "Glass, Chemical Decomposition of."

The best European deposits are at Fontainebleau near Paris, and Lippe, Saxony. These contain over 99·7 per cent of silica and less than 0·03 per cent of iron oxide; they are remarkably uniform in composition and grain size. Next in order of merit come the Belgian sands, notably at Epinal (below 0·05 per cent of iron oxide). Fair quality Dutch sands are also found.

Some of the British sands have a low iron content, but, in general, the sand is not so uniform in quality as that of Fontainebleau. Good deposits are found at Lynn, Aylesbury, Muckish Mount, Huttons Ambo, Burythorpe, and other places.[1]

For the manufacture of bottles sands rich in alumina are useful.

A number of good deposits of glass sands are found in America, notably in Illinois and West Virginia.

§ (5) ALKALIS.—(i.) *Sodium oxide* is generally derived from sodium carbonate ("soda ash") obtained from the Le Blanc or ammonia soda process, or sodium sulphate ("salt cake") obtained from the Le Blanc process. Until recently, owing to its relatively low cost, salt cake was largely used as a source of soda, but its use is attended by various disadvantages. It is usually contaminated with chlorides, and occasionally free sulphuric acid, and it is not easy to remove all the sulphates from the glass during manufacture. The chlorides and sulphates tend to produce a milky appearance in the glass. A high temperature is necessary to decompose the salt cake, the decomposition being assisted by the addition of carbon to the batch; the presence of carbon, however, makes the use of a "decoloriser" extremely difficult. Another disadvantage in the use of salt cake is that the walls of the furnace are seriously corroded along the "flux line" (at the surface of the glass) by the alkali which floats on the surface of the half-melted batch.

Soda ash, though formerly expensive, is now cheaper than salt cake and is becoming more popular owing to the inconveniences inherent in the use of salt cake. With soda ash as the source of alkali, however, it is difficult to melt glasses rich in silica and lime without the formation of a scum rich in silica which floats on the surface of the glass. A common practice is to use soda ash with the addition of the minimum amount of salt cake to prevent the formation of the scum.

Small quantities of soda are also introduced in the form of sodium nitrate for the purpose of oxidising any organic matter, ferrous iron, etc., that may be present in the glass, and of securing the oxidising conditions necessary for melting batches containing lead, which is easily reduced.

(ii.) *Potash* is universally added in the form of potassium carbonate ("pearl ash"). This can be obtained commercially in a state of reasonable purity. Pearl ash is highly hygroscopic, and careful analytical control is necessary if potash glasses are required to be of constant composition. Potassium nitrate (saltpetre) is also used in small quantities.

(iii.) *Felspar* (K_2O, $Al_2O_36SiO_2$) is occasionally used as a source of alkali where the presence of alumina is not considered disadvantageous, and, for cheap glass-ware, bottles, etc., rocks such as granite and basalt are sometimes used.

§ (6) LIME is usually added in the form of ground chalk, lime spar, or limestone rocks. In glasses containing magnesia dolomite provides a convenient source of lime and magnesia. Where greater purity is required, precipitated calcium carbonate can be obtained.

§ (7) BARYTA.—The cheapest source of barium oxide is the mineral witherite, $BaCO_3$, which can be obtained in sufficient purity for the manufacture of some glasses, pressed glass, etc. For optical glasses chemically prepared precipitated $BaCO_3$ or $Ba(NO_3)_2$ is used.

§ (8) LEAD OXIDE is usually added in the form of red lead (a mixture of PbO and Pb_2O_3, corresponding roughly to Pb_3O_4). Red lead varies somewhat in composition, free lead and lead sulphate are sometimes present, and the moisture content is variable. It is desirable that the maximum amount of Pb_2O_3 be present to facilitate oxidising conditions. Litharge, PbO, is sometimes used in the batch, but, owing to its low oxygen content, it is not desirable.

§ (9) MAGNESIA.—The chief sources of magnesia are dolomite (calcium and magnesium carbonates) and magnesite ($MgCO_3$). These minerals are obtainable in reasonable purity and are used in considerable quantities. For special purposes precipitated magnesia is used.

§ (10) ALUMINA.—For the best glasses the sources of alumina are plentiful, china clay, felspar, etc. There is some difficulty, however, in obtaining a cheap form of alumina for the manufacture of common glass bottles, where the low cost of the batch makes the addition of felspar too expensive. This is unfortunate, since the presence of alumina in glass increases the durability both mechanically and chemically.

§ (11) BORIC OXIDE (B_2O_3), an ingredient of many optical and resistance glasses, is introduced either as boric acid (H_3BO_3) or as borax.

§ (12) MISCELLANEOUS INGREDIENTS.—For the manufacture of opal glasses fluorspar (CaF_2) and cryolite (AlF_33NaF), both natural and artificial, are used. Calcium phosphate is also used for this purpose.

Other materials required in glass-making are used in the form of chemically prepared com-

[1] P. G. H. Boswell, "British Glass-making Sands"; C. J. Peddle, "British Glass-making Sands," *Jour. Soc. Glass Tech.* i. 27.

pounds, since they are only required in small quantities, and the cost is therefore unimportant.

§ (13) Arsenic.—White arsenic or arsenious oxide (As_2O_3) is the usual source of arsenic for the glass industry. Arsenic is introduced into both oxidising and reducing batches with apparently beneficial results. It appears to accelerate the removal of bubbles from the glass and, also, to assist "decolorisers" in their action.

It is probable that under oxidising conditions the arsenic is oxidised at low temperatures to As_2O_5, and at high temperatures this oxide tends to decompose, forming As_2O_3 and liberating oxygen which sweeps out the small bubbles from the glass and also oxidises the manganese and iron, giving the condition required for successful "decolorising." Under reducing conditions metallic selenium is generally used as a decoloriser. A part of the arsenic is reduced to metallic arsenic, which is volatile, and thus sweeps out the bubbles with a reducing vapour which tends to keep the selenium from becoming oxidised.

Analyses of glasses made from batches containing nitrates show that from 10 per cent to 20 per cent of the arsenic may be lost during melting, and, of the arsenic remaining in the glass, from 80 per cent to 90 per cent is in the pentavalent condition. Similar glasses melted without nitrates lose from 30 per cent to 40 per cent of the arsenic during the melting process, and only 60 per cent of the remaining arsenic is in the pentavalent condition.

V. The Manufacture of Glass

It is not possible here to describe in any detail the manufacture of glass, since the process varies very considerably with the type of glass made and the use for which it is intended. The general outlines are given below.

§ (14) The Preparation of the Batch.—The term "batch" is used to denote the ingredients of the glass mixed in the required proportions and ready for melting. The success of the melting operation and the quality of the glass obtained depends very largely on the care taken in the preparation of the batch. Many manufacturers do not sufficiently realise the importance of this side of the glass works. The raw materials should be well ground. Even grading is as important as the actual fineness of division. The sand, and occasionally other ingredients, are dried; in some works the sand is roasted to a red heat before use. The various constituents are then weighed out and fed into a mechanical mixing machine, which delivers the mixed, sieved batch into a container in which it is conveyed to the furnace. In some up-to-date, well-organised works elaborate plant is used to carry out these operations, and the weighing, mixing, and conveying of the batch is effected automatically. Many works are to be found, however, where the batch composition is estimated by volume—in some cases the shovel being the unit of volume—and the mixing carried out by hand.

§ (15) Furnaces and Melting Process.—Two entirely different methods of melting are employed according as the glass is contained in crucibles (or "pots") or in a tank.

(i.) *Melting in Pot Furnaces.*—The pots used to contain the glass during melting are of two kinds, open and closed. The open pot is a cup-shaped vessel, anything up to 5 ft. in diameter and 5 ft. high (or even larger), containing, when full, from 2 cwt. to 2 or even 3 tons of glass.

The closed pots are similar to the open ones except that they are fitted with a dome or hood which has an opening on one side through which the batch is charged, and the finished glass subsequently gathered. This hood, which projects through the wall of the furnace, serves to protect the glass from contamination by the furnace gases.

One furnace may contain several pots, from 2 up to 20 or more. The furnaces are usually fired with producer gas, although direct coal-fired furnaces are still found in some parts.

Great care has to be exercised in the manufacture and subsequent handling of the pots. These are made from carefully selected fire-clays (see "Glass-house Refractories," Vol. V.). The raw clays are mixed with a quantity of burnt clay, to reduce the shrinkage on firing, moistened, and allowed to mature for some months. The pot is then built up slowly by hand. The pot-maker first moulds a disc to form the bottom of the pot and then proceeds to build up the sides; not more than 6 in. or so of wall can be added at a time, since the plastic clay is not strong enough to support the weight of the whole side. The pot has, therefore, to be built in stages with an interval of a day or two between each stage. In some works plaster or wooden moulds are used to support the pots during their manufacture. Open pots are occasionally made by the casting process (see "Glass-house Refractories," Vol. V.). When the pot is completed it is allowed to dry slowly. The temperature of the drying-room is carefully regulated and precautions are taken to prevent draughts. When the pots are dry, a process which generally takes some months, they are removed from the drying-room when required for use and placed in a small kiln (the "pot arch"), in which they are brought slowly up to a red heat. Very great caution is required in the initial heating, and the correct rate of rise of temperature, which varies with different clays, must be maintained carefully. The temperature to which the pot is taken in the pot arch varies considerably in different

works, and frequently 900° C. is not exceeded. It is better practice, as a rule, to run to considerably higher temperatures. When the pot arch is at its maximum temperature, the pot is taken out and transferred as rapidly as possible to the melting furnace. For this purpose the brick work in the front of the melting furnace is removed, leaving a hole sufficiently large to admit the pot. When the pot is in position in the melting furnace, the wall is bricked up and the temperature raised. In the best practice the pot is heated to a temperature considerably higher than that required for the melting of the glass, since the life of the pot is prolonged if it is well "vitrified" before the glass is charged into it. In many works, however, the pot is not subjected to this preliminary heating and the glass is introduced as soon as the pot reaches the melting temperature of the glass, with the result that much unnecessary corrosion of the pot takes place.

The raw materials are mixed with a quantity of previously-melted glass of the same composition (cullet). This greatly facilitates the melting process. The pot, when ready for filling, is first glazed with cullet to protect it from the direct contact with the raw materials, particularly the alkali, which has a powerful fluxing action on the clay. The pot is then filled with the raw material and cullet. Since there is a considerable diminution in bulk when fusion takes place, several fillings are required to yield a pot full of molten glass. The method of filling varies somewhat in different works. Sometimes great care is taken to prevent the batch from touching and chilling the walls of the pot. In this case the batch is piled up in the form of a cone and as many as six or more fillings are made. In other works the batch is shovelled on quickly, two fillings, or occasionally, with lead batches, one filling, being considered sufficient.

The contact of the damp material with the walls of the pot and the consequent chilling of the inside frequently gives rise to small cracks in the pot walls (known as "batch cracks"). These increase in size at each filling and often cause the ultimate failure of the pot.

The reactions which occur during fusion are complex and imperfectly understood. The order in which the various changes take place is roughly as follows. The moisture and combined water is expelled from the batch:

Alkalis in the form of nitrates melt	320°
Magnesium carbonate decomposes	350°
Red lead decomposes, forming litharge	500°
Boric oxide melts	577°
Barium carbonate melts	795°
Potassium carbonate decomposes	810°
Calcium carbonate decomposes	825°
Sodium carbonate melts	849°
Litharge melts	877°
Potassium oxide melts	880°

Complex eutectics are formed, in which the oxides having high melting-points are slowly dissolved. The ease with which the glass melts depends largely on the viscosity of these eutectics. Owing to the low diffusivity of the mixture a very large temperature gradient is set up in the batch, and, in consequence, all the possible physical and chemical changes occur simultaneously in different parts of the pot, and many bubbles of gas, water, carbon dioxide, etc., are entrapped in the molten glass. The raw materials of the glass are eventually dissolved, and the bubbles of gas rise slowly to the surface at a rate depending on the square of their diameter, the viscosity and density of the glass. By a judicious choice of raw material and temperature of melting it is possible to arrange that only large bubbles remain in the glass at the end of the fusion, and these rise to the surface in a few hours and burst. If for any reason small bubbles (technically known as seed) are formed, the glass-maker frequently has to introduce large bubbles to sweep the small ones to the surface. For this purpose it is common practice to plunge a potato, on the end of an iron rod, to the bottom of the pot. The moisture and organic matter, which is given off violently, effectually removes the small seed. Nitrates and arsenic are sometimes used for this purpose.

To facilitate the removal of the bubbles (a process commonly known as "fining"), the temperature of the furnace is raised. With the diminution in the viscosity of the glass, the corrosion of the pot becomes severe. The glass in contact with the pot walls dissolves the material of the pot, the resulting glass having a density different from the mean density of the melt. Currents are therefore set up and, since the density of most glasses is greater than that of the pot used, these currents, caused by the solution of the pot in the glass, flow up the sides of the pot. As the glass rises, it becomes richer in pot material, and, consequently, less active as a solvent, with the result that the attack of the glass on the walls of the pot is greatest at the bottom and a taper is produced in the pot.[1]

There are other factors which affect the rate of corrosion (apart from the imperfections in the pot itself), *i.e.* the temperature distribution in the furnace, since it is possible for large convection currents to be set up if the bottom of the pot is heated locally. (See "Glass-house Refractories," Vol. V., for further details on the properties and behaviour of glass-house pots.) The presence of these convection currents is important, since the rate of solution of the

[1] Coad-Pryor, "Notes on Pot Attack," *Jour. Soc. Glass Tech.*, 1918, ii.; and Rosenhain, "Some Phenomena of Pot Attack," *Jour. Soc. Glass Tech.*, 1919, iii.

pot in the glass depends largely on the rate at which the glass rich in dissolved pot material can be removed from the vicinity of the pot and fresh glass brought in contact with it. The presence in the glass of an excess of silica and alumina derived from the pot increases the viscosity greatly, giving rise to a corresponding diminution in the rate of diffusion, and, but for this formation of currents, the pot might reasonably be expected to last for years instead of months as is actually the case.

The formation of sillimanite crystals (Al_2O_3, SiO_2) in the pot appears to exert a marked beneficial influence on the rate of solution.

During the "fining" process samples (proofs) are taken from the pot on an iron rod and, from these samples, the expert glass-maker can decide when the "metal" is ready for working or gathering. When the glass is fine (free from bubbles and raw material) the temperature is lowered until the viscosity suitable for gathering is reached. The glass is still far from being homogeneous, the lower portion being considerably denser than the upper, and the parts near the sides being richer in silica and alumina (derived from the pot) than the central portion of the pot. However, for all purposes except for optical instruments, the glass is considered good enough for use.

(ii.) *Melting in Tank Furnaces.*—For the production of cheaper kinds of glass where mass production is essential, "tanks" are invariably employed for melting purposes. A tank consists of a fire-brick chamber which may vary considerably in size according to the output required, their capacity ranging from two to over a thousand tons of glass. The ratio of length to breadth is roughly three to two, average dimensions being from 12 to 18 ft. wide, 24 to 36 ft. long, and from 30 to 44 in. deep. The bottom and sides of the tank are usually constructed of firebrick or refractory stone and the crown of silica brick. The gas and air enter by ports arranged along the side of the tank and the combustion takes place over the surface of the glass, the products of combustion passing out on the opposite side by flues leading to the regenerators, which are usually situated below the tank.

The dimensions of the tank and the distribution of temperature are so contrived that the raw materials are fed on at one end of the tank, melt and flow slowly down the tank, and, by the time the glass reaches the other end, it is fine and ready for working.

This procedure is more economical than melting in pots, since the glass is exposed to the direct heat of the flame. The process is continuous and the glass may be worked day and night without cessation. The refractory materials are more durable, since, in the places where they are in contact with the glass, they are always cooler than the glass, whereas, in a pot furnace, the heat has to be transmitted through the pot walls.

Frequently the most vulnerable parts of the tank are water- or air-cooled on the outside and, consequently, the tank is practically lined with a layer of glass too viscous to attack the refractories at an appreciable rate. In some works the whole of the furnace walls are water-cooled, and, in such furnaces, the rate of solution of the refractories is extremely slow.

No thermal convection currents are found in the glass in a tank, since the heating is from above, and we find, therefore, that the tank blocks wear away first near the surface of the glass, the attack on the blocks diminishing with the temperature as the depth below the surface of the glass increases. The blocks forming the bottom of the tank may last for some years, but the side walls, near the surface, seldom last for more than a year, often less.

In a furnace melting flint (colourless) glass the serious corrosion extends to a depth of perhaps 18 inches or 2 feet; with deeply coloured glass, an amber or dark green, the attack of the glass on the sides of the tank 8 or 10 inches below the glass level is slight.

During the melting process, in both pots and tanks, it is frequently found that a scum is formed on the surface of the glass. In pots, the glass is skimmed before the gathering of the glass is commenced; in tank furnaces, to meet this difficulty a wall or bridge is generally built across the tank, shutting off the working end from the melting. This bridge, which projects a few inches above the surface of the glass, retains the scum and the surface layer of the glass generally rich in silica, and allows the clear glass below to flow into the working end through a hole in the bridge near the bottom of the tank. The attack of the glass on the bridge is severe, and it is generally necessary to cool it by building it hollow and maintaining a circulation of air or steam in it. The cooling of the orifice through which the glass flows presents special difficulties, since severe chilling will cause obstruction due to high viscosity of the glass. Various devices have been adopted to separate the working end from the melting end of the furnace, since the temperature suitable for gathering the glass is lower than that required for melting. Thus the bridge is sometimes continued right up to the crown of the furnace, the part above the glass level being built in chequer work to cut down the radiation from the melting end. Again, a separate chamber connected to the melting tank by means of a clay syphon has been used. Sometimes small covered pots (generally known as "potettes") are built into the working end of the furnace with their hoods projecting through the furnace walls in a manner similar

to pot furnace practice. These potettes are drilled near the bottom, below the surface level of the glass, to allow the glass from the tank to flow into them. When working with potettes a bridge in the tank for skimming purposes can be dispensed with.

§ (16) DEFECTS. (i.) *Stones.*—If the process of manufacture outlined above has been carried out successfully, glass is obtained in the pot or working end of the tank free from bubbles and undissolved material and reasonably uniform in composition. There are various defects, however, which are frequently found which have arisen through accidental causes or through carelessness on the part of the operators. Of these perhaps the most common is the disease known as "stones." These consist of solid particles which have not been taken into solution in the glass, and may be caused by drips from the crown of the furnace or particles of refractories from the walls which have become undermined and detached. Stones may also be formed if parts of the furnace have been allowed to become too cool, when "devitrification" or crystallisation of the glass may result. These crystals, when once formed, are only dissolved with difficulty and, in consequence, give rise to stones. One of the most prolific sources of stones is the inclusion in the batch of coarse particles of raw material, such as limestone.

(ii.) *Striae.*—Another common defect is the presence of striae or "cords," which consist of veins of glass of a different refractive index from that of the rest of the glass. Strictly speaking, striae are always present in any but the best optical glass, although, for most purposes, they are sufficiently small to pass unnoticed. They are produced by imperfect mixing of the batch or contamination of the glass by the walls of the pot or tank. Stones always give rise to striae, which persist in the glass after the stones have disappeared. Striae are also found if the furnace has been allowed to run cold during the melting, when the high viscosity has hindered the diffusion of the constituents.

(iii.) *Seed.*—The presence of fine bubbles, or "seed" as they are called, in the finished glass is a defect sometimes found. Bubbles, which should have been removed during the fining process, are sometimes found in the glass owing to inadequate control of the temperature during founding or to an attempt to increase the rate of production beyond the working capacity of the furnace.

Seed may appear, however, at a later stage, after the glass has once been fine. This phenomenon is common in amber glasses where the colouring has been produced by carbon, probably in colloidal solution. Under certain conditions a rise in temperature or mechanical agitation may cause an evolution of gas. A similar effect is said to occur in cobalt blue glasses. Again, with some glasses, certain gases appear to be more soluble at high than at low temperatures, and on cooling bubbles are evolved. By successively cooling and heating the glass for a number of times the quantity of gas in solution can be diminished, and a marked improvement is observed in the tendency of the glass to devitrification. It has been shown that glass, under normal conditions, can hold in solution considerable quantities of gas. Allen and Zies obtained from 6·5 grms. of glass 6·5 c.c. of gas having the composition: oxygen, 64·2 per cent; carbon dioxide, 24·2 per cent; carbon monoxide, 3·5 per cent; hydrogen, 3·9 per cent; nitrogen, 4·1 per cent; and instances have been given where glass on cooling has liberated many times its own volume of gas. All glass, as usually made, contains a large amount of dissolved water vapour, far in excess of other gases found in glass. This is evolved copiously on heating the glass to 300°-400° C. or over, under reduced pressure.

So-called "vacuum bubbles" are formed where the surface of the glass is chilled rapidly, forming a surface crust, while the centre is still hot. As the interior of the glass shrinks a high state of tension is produced and minute bubbles may swell to a large size.

(iv.) *Colour.*—A frequent source of trouble to the glass-maker is the control of the colour of the glass. Iron is always present in the raw materials and refractories, and, when present in an excessive amount, produces a colour which varies from a yellowish green to blue according to the state of oxidation of the iron and the composition of the glass. The remedy for this is, of course, to use materials as pure as possible, but, for most commercial purposes, raw materials free from iron are much too costly. To mask the iron green, a colouring agent which gives the complementary colour, pink, is added to the batch. Of these "decolorisers," the most usual reagents are manganese (used with batches worked under oxidising conditions) and selenium (for reducing conditions), often in conjunction with small quantities of chromium, nickel, and cobalt. The colouring effect due to the presence of iron is less when the iron is in the trivalent condition, and hence the presence of nitrates in the batch improves the colour, since oxidising conditions are then attained. It is not usual to attempt to decolorise glasses containing more than 0·2 per cent of iron calculated as Fe_2O_3; a higher content of iron gives a glass which, when decolorised, has a perceptible grey tint.

A good "flint" soda-lime glass may contain

about 0·11 per cent Fe_2O_3, and a lead flint glass for cut glass table ware may contain 0·03 per cent Fe_2O_3 or even less.

Arsenic is also introduced into the glass to facilitate the action of decolorisers. The chemistry of this action presents some difficulties. (See "Raw Materials," "Arsenic.")

§ (17) Optical Glass.—The standard of quality required in glass for optical purposes makes the manufacture of optical glass a special process. The glass requires to be carefully stirred to produce the necessary homogeneity, and, even then, it is considered a good yield to obtain 20 per cent of good glass out of the pot. For this purpose a clay stirrer is introduced into the pot as soon as the glass is fine, or sometimes at an earlier stage, and the glass is slowly stirred. This operation is carried out mechanically, the clay stirrer being manipulated by means of an iron rod. It is necessary to stir as the temperature falls, until it is no longer possible to continue owing to the high viscosity. The stirrer is then either removed or allowed to remain in the glass near the side of the pot. The rate of stirring permissible varies within narrow limits, and has to be regulated carefully as the glass cools. At the end of the stirring operation the pot is removed from the furnace as quickly as possible and allowed to cool rapidly. With some glasses which tend to devitrify it is necessary to chill the glass by spraying water on to the outside of the pot. When sufficiently cold the pot is broken up and the glass carefully examined, those pieces which are considered free from striae and other defects being reheated to the softening point of the glass and moulded into rectangular blocks or lenses. The blocks are then annealed (see "Annealing," § (10)), and finally ground and subjected to further inspection.[1]

§ (18) Manipulation.—The general principles of glass-blowing are well known and have been described frequently.[2]

A brief outline of the processes in use at the present day is given below.

(i.) *Hollow Glass-ware.*—Although glass made by mechanical means has displaced in many cases the hand-blown articles, it is difficult to obtain by machinery the excellent finish which can be produced by a skilled glass-blower. The automatic production of unsymmetrical shapes, *i.e.* water jugs with handles and spouts, and complicated pieces such as wine-glasses, presents serious difficulties, and, at present at any rate, the supremacy of hand-blowing for such ware is unchallenged.

The blowing of a large glass article is generally carried out by a team of men or boys, who constitute a "chair." As a rule, two men and two boys make up a chair, although, for some work, only three persons are employed. In some works there is a tendency to increase the number of men to a chair, since, by so doing, the work becomes more highly specialised, and each man, in consequence, more efficient at his small contribution, and the highly skilled labour economised.

The glass-blower's equipment is simple, the most important item being the stool on which he sits. This consists essentially of a small wooden bench provided with arms on each side which project in front and behind him.

The "gatherer" first takes the blowing-iron, an iron tube about 5 ft. in length, which has previously been heated, and dips the end in the glass with a circular motion, withdrawing it with a ball of glass adhering to it. He then rolls the glass on an iron plate (the "marver") in such a way as to obtain a symmetrical shape in the glass on the blowing-iron. During this operation the glass has become chilled and stiff to work. He then dips it again in the pot, gathering more glass and marvering as before. This process is repeated until the required weight of glass has been collected on to the blowing-iron. From time to time the gatherer blows gently down the tube, expanding the glass on the iron into a pear-shaped mass. He then hands the iron to the glass-blower, the head of the "chair," who sits on his bench and places the blowing-iron across his knees, resting it on the arms of the chair. Since the glass is mobile during the blowing operation it is necessary to keep the iron constantly revolving to prevent the gathering from losing its symmetrical shape. The blower, by rolling the iron along the arms of the chair with one hand, is able to effect this, and, holding his tools in the other hand, he can shape the article as though the iron were in a lathe.

The glass-blower's outfit consists of a pair of shears or scissors, a palette or "battledore" (a small wooden board with a handle), a pair of tongs, calipers, a foot-rule, and an iron bar or file.

In the process of blowing hollow ware such as a jug or decanter, he rests the iron on one arm of the chair and blows until the desired size of bulb is obtained. He then shapes the bottom and sides of the jug by revolving it on the two arms, as indicated above, with one hand, and pressing the glass into the required shape with the palette, and checking the dimensions with the calipers.

[1] Series of papers issued from Geophysical Laboratory, Washington, on optical glass; Rosenhain, *Glass Manufacture*, ch. xiv.; Rosenhain, *Optical Glass: Cantor Lecture to Royal Soc. Arts*.

[2] Rosenhain, *Glass Manufacture*; Powell and Rosenhain, articles on Glass in the *Encyclopædia Britannica*, 9th and 11th editions; Powell and Chance, *Principles of Glass-making*; P. Marson, *Glass*.

When he has trued the bottom of the jug, one of his assistants takes an iron rod—the punty (also known as the pontee, ponty, or puntee)—gathers a small amount of glass on the end of it, and then presses this on to the centre of the bottom of the jug. The glass on the punty adheres to the jug; at the same time the blower touches the glass at the end of his iron with the cold iron bar, and the jug breaks off the blowing-iron and remains attached to the punty. The assistant then reheats the jug, which has by this time become set, in a small furnace, the "glory hole," and when it has softened again returns it to the glass-blower, who revolves it rapidly on the bench. The centrifugal force causes the hole at the top of the jug (where it was originally attached to the blowing-iron) to open out. The blower trims the end with the shears, and then, rotating it, forms it into the desired shape with the help of his tongs. In this manner, with modifications suited to the type of article being made, all kinds of hollow glass-ware can be blown. In many cases the process of blowing may be simplified by the use of an iron mould made in two parts, with an opening at the top, so constructed that the blower can open and close the mould by means of a lever operated by his foot. He blows a pear-shaped bulb which he holds vertically for a second or so until an elongated neck is formed; he then introduces the bulb into the open mould with the neck protruding through the hole in the top. The mould is then closed, and the bulb blown up until it fills the mould. When the glass has set, the mould is opened and the finished article is removed and cracked off the blowing-iron. During the blowing up in the mould the blower spins his blowing-iron, rotating the glass in contact with the mould, and by this means prevents a mark from the joint in the mould being imprinted on the glass. Bottles, etc., not spun in this way always show a mould mark.

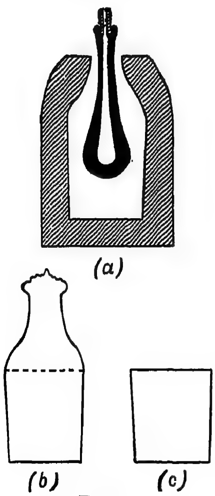


FIG. 1.

Fig. 1 (*a*) shows a diagram of a section through a mould for blowing a laboratory beaker or a tumbler, and (*b*) and (*c*) show the successive stages; (*b*) shows the tumbler as it leaves the mould. After annealing it is placed on a revolving disc, so arranged that a pointed flame plays on the glass (along the dotted line) as it revolves; the glass is then removed from the flame and touched with a cold, pointed, steel rod at any spot on the ring heated by the flame. If the glass has been well annealed, it will crack evenly along this line, leaving the tumbler as in *Fig.* 1 (*c*). It is then placed on a revolving disc or clamped to a rod which can be made to rotate horizontally, and the rim is heated until the glass softens and a smooth edge is produced by the surface tension. When the lip requires flanging the heating is carried further and the edge turned over with a tool to the amount required as the glass revolves. This operation introduces severe local strain round the rim of the glass, which has, in consequence, to be re-annealed.

(ii.) *Tube and Rod.*—In the manufacture of tubing, the gatherer prepares a large mass of glass on his blowing-iron, the size (as a rule from 5 to 20 lbs.) depending on the diameter and thickness of wall required in the tube. The gathering is carefully marvered and reheated in the glory hole. The blower then takes the blowing-iron and allows the glass to elongate by holding the iron vertically and giving it a gentle swing. His assistant then holds the punty in a convenient position, and the blower, by a dextrous movement, attaches the end of the gathering to the punty. The two men then move apart at a rate governed by the dimensions of the tube they are making. For small tubes it is necessary to run, but for tubing of medium diameter a slow walk is usual. With heavy tubing which tends to sag excessively it is necessary to chill it during the drawing with an air-blast, or, more usually, by fanning it. Solid rod is made in precisely the same manner, except that the gathering is made with an iron rod instead of a tube.

(iii.) *Glass Wool.*—By drawing out glass sufficiently quickly very fine fibres can be produced: thus, glass wool is made by winding glass drawn from the gathering on to a rapidly rotating drum.

(iv.) *Sheet Glass.*—A gathering of glass is made in the usual manner. The blower stands on a raised platform, so situated as to give him room to swing his iron vertically below him. He is usually provided with mechanical devices to assist the blowing operation, *i.e.* a flexible tube to enable him to use compressed air for blowing, and arrangements to relieve him of part of the weight of the gathering. He is also conveniently near a glory hole, for reheating during the blowing.

By means of specially shaped blocks and tools, the gatherer marvers a short, wide-bore cylinder having a thickened end, as in *Fig.* 2. By alternately heating the end of the cylinder and swinging it in a vertical plane, a cylinder,

about 4 ft. long or more, of reasonably uniform bore and thickness of wall, can be made. The end of the cylinder is then opened by bringing a small gathering of glass on the punty in contact with it; the glass of the cylinder softens at the point of contact, and a small hole is cut out with the shears. The end is then reheated, and, by rotating the cylinder, the hole becomes enlarged by the centrifugal force until its diameter is the same as that of the cylinder. The cylinder is then laid on a wooden stand, detached from the blowing-iron and the ends cracked off, generally by passing a thread of hot glass round it and touching it with a cold iron tool. A longitudinal scratch is then made from end to end on the inside surface, and a slight tap causes a crack to run down the side of the cylinder along the scratch. It is conveyed to the flattening kiln and placed on a smooth slab; the temperature is raised to the softening point of the glass, and the cylinder slowly opens out until it becomes approximately a plane sheet. The surface is rubbed with a wooden block to secure a smooth surface and intimate contact with the slab on which the sheet rests, and the finished sheet is then passed down the annealing lehr which adjoins the flattening kiln.

FIG. 2.

(v.) *Plate Glass.*—The manufacture of large sizes of the best plate glass is an expensive and wasteful process. Economical methods of making small-sized plates have been developed, but for very large sheets the casting and rolling process has not been supplanted.

A large mass of glass is poured on to an iron table; this is effected either by ladling the glass from a tank or pot into the casting pot in which it is carried to the table, or by removing the melting pot itself from the furnace. In either case the operation is carried out largely mechanically. The glass is then rolled with an iron roller running on rails at the edges of the casting table, the height of these rails determining the thickness of the plate. The roller is propelled mechanically, the width of the plate being controlled by guides which the roller pushes in front of it as it progresses down the table. After rolling, the sheets are annealed, either in a continuous lehr along which the sheets travel slowly, or in a kiln which is heated to receive the sheet and then allowed to cool down slowly, the sheet taking several days to reach a temperature at which it can be handled.

For certain purposes (roofing, etc.) sheet made in this way is sold without further treatment, but the surface is poor and the glass quite unfit for the purposes for which the best plate is required.

The plates have then to be ground to a true surface and polished. For this purpose they are clamped to a circular iron table, about 30 ft. in diameter, which is rotated, and the surface of the plate is rubbed down with iron slabs which are also rotated, eccentrically to the table. (The plates are usually set in plaster of Paris to ensure that the under surface is in intimate contact with the surface of the table.) The iron rubbers are fed with abrasives of successively finer grades (starting with coarse sand), until, finally, a fine ground surface is produced in the glass. During this process a large amount of glass (up to 30 per cent) is ground away and lost. The sheet is then polished, the iron "rubbers" being replaced by wooden blocks covered with cloth or felt and fed with rouge. If the previous grinding process has been carried out successfully, a brilliant surface is obtained with rouge in a short time—two hours or so. The sheet has then to be turned over and the whole process of grinding and polishing repeated on the other side. It is then washed and inspected, and cut to the required sizes, in such a manner as to exclude any defects which may be present.

Sheet known as figured rolled glass is made in large quantities. This is rolled in a somewhat different manner, four rollers being used. Instead of moving rollers such as are used in the method described above, these rollers are mounted in pairs and rotate about stationary axes, the glass being drawn between the two rollers. The sheet is formed by the first two rollers, and, passing through the second pair, receives the imprint of a pattern which has been cut on one of these rollers. Rolled sheet with simple patterns on it, such as parallel lines, etc., is also made by cutting the surface of a casting table to the desired pattern and rolling with a single roller as described above. The pattern is only pressed on one side, since the method adopted in cutting up the sheets, *i.e.* scratching with a diamond point and then breaking along the scratch, cannot be applied conveniently if the surface scratched is not plane.

Where plate glass having a curved surface is required (such as that sometimes used in large shop windows), the plate is heated on a mould having the desired curvature; the glass softens and takes the shape of the mould. Great care is necessary to keep the surface clean and free from defects during this operation, since any dust, etc., settling on the surface gets burnt in when the glass softens in bending. It is generally necessary to repolish parts of the surface, by hand, after the bending process.

(vi.) *Pressed Glass.*—Many kinds of cheap glass-ware, principally shallow vessels or solid objects, are made in presses, the glass being deposited in a mould and pressed into shape with a plunger.

A number of different types of machines have been developed for this purpose, some being operated by hand and others being automatic. Some machines combine the operations of pressing and blowing. Machines of this description are in common use for the manufacture of such articles as jam jars, heavy wide-mouth bottles of all kinds, tumblers, etc. The glass is fed, either by hand or by an automatic feeding device, into the first mould, the "parison" mould, and a plunger presses the glass into a symmetrical shape, this part of the process corresponding to the marvering in a hand-blown article. The parison is considerably smaller than the finishing mould except at the neck, where the glass is pressed at once to the full finished size and chilled, thus affording a means of conveying the half-formed jar to the finishing mould. The jar is taken from the parison mould and transferred (holding it by the chilled neck, the lower part of the jar being still plastic), either by hand or automatically, to the finishing mould, in which it is blown up to the full size. The finishing mould is then opened (it is generally made in two parts), and the jar is conveyed to the lehr.

The tendency of modern practice is to develop automatic processes, to increase and standardise production, and to cut down labour expenses. Thus in an up-to-date bottle works the entire process, batch-weighing, mixing, conveying, and charging into the furnace, is automatic, as is also the blowing of the bottles and the conveying of the bottles to the annealing kilns (known as "lehrs").

There are many types of machines in use for the manufacture of hollow ware, bottles, jars, electric light bulbs, etc., some entirely automatic and others semi-automatic, *i.e.* requiring the assistance of a "gatherer," who gathers the glass from the furnace by means of an iron rod or tube and keeps the machine supplied.

Of the fully automatic machines the Owens machine is perhaps the best known. This remarkable machine sucks the glass from the surface of a specially constructed tank into a mould (the parison mould), which it then withdraws from the tank, filled with glass, trimming off the glass adhering underneath the mould with a knife. The neck of the bottle is chilled and the parison mould opens, leaving the half-formed bottle hanging by the neck. The finishing mould then closes round the bottle, which is at the same time blown out to its full size. This mould then opens and the bottle is thrown out into a chute which conveys it to the lehr. Each mould is mounted on one of a number of rotating arms, and, by this means, gathers the glass as it passes the tank and discharges its bottle just before it completes one revolution or later, according to the type of machine. In order to present a continually fresh surface of the glass to the machine the hearth of the tank is revolved. The glass in the revolving tank is maintained at the same level, being fed continuously from the melting tank. The output from these machines is enormous, one 15-arm machine making up to 1000 gross of small bottles per day.[1]

There are many machines which are fed with a flow device, the glass flowing from the tank through a heated clay trough into the parison moulds. The rate of flow can be regulated by means of clay stoppers in the trough.

The manufacture of light articles, such as electric bulbs, is now extensively carried out automatically, the Westlake machine being perhaps the best known for this purpose. This machine gathers the glass, by means of a small cup, from a crucible into which the glass is ladled from the melting furnace. A 12-arm Westlake machine will produce up to 100,000 bulbs per day.[1]

Processes are also in operation for the automatic manufacture of window-glass and glass-tubing. In the window-glass process the glass flows or is ladled from the melting tank into a trough or basin, and a suitable "bait" is lowered on to the surface of the glass. The glass adheres to the bait, which is then slowly raised, drawing a sheet or tube (according to the shape of the bait) of glass from the trough. If the temperature is carefully controlled and the correct rate of drawing maintained, the process is continuous, the glass being drawn over rollers and through the lehrs to the sorting-house. In the tube process, when the required length has been drawn, the bait is raised quickly and the glass detached from the basin. The ends of the cylinder are then cracked off by passing a wire round it, and it is then scratched longitudinally and opened out in a manner similar to the familiar hand process of making window-glass.

Tank furnaces are now employed for melting glasses which, a few years ago, were invariably melted in pot furnaces; for example, glass for electric-light bulbs, tumblers, chemical ware, tubing, table-ware, opal glass; and it is claimed that lead glasses have been successfully melted in tanks.

The old rule-of-thumb methods are gradually being displaced in favour of more scientific

[1] For a description of some of the principal glass-making machines in use see *Journ. Soc. Glass Tech.*, 1917, p. 203; 1918, p. 19; 1919, p. 182.

control, and we find, for instance, pyrometers and polariscopes in common use. However, in spite of the rapid advance in the engineering aspect of glass technology, there will always be a demand for the highly-skilled glass-blower for the manufacture of artistic table-ware, cut glass, and so on.

§ (19) ANNEALING.—The viscosity of glass rises very rapidly with a fall in temperature. It is estimated that up to, say, 650° the viscosity is doubled for every 6° to 12° fall in temperature.

(i.) *Formation of Strain.*—When a block of glass is cooling uniformly from all sides there will be a temperature gradient from the outside to the inside which will depend on the thermal conductivity of the glass and the rate of cooling. Since the thermal conductivity of glass is low, very large temperature gradients are found if the heat be removed rapidly from the surface. As long as the glass is mobile no strain will be introduced. When, however, the viscosity of the outer layer becomes sufficiently high to resist the compression due to the shrinking of the inner layers, permanent strain may be produced. Let us suppose that when this viscosity is reached the outer layer is 600°. The centre of the block is then at, say, 650°. It is clear therefore that, if a is the mean linear coefficient of expansion of the glass, the inside layers will tend to shrink approximately $650a$ per unit length when cooled to 0°, whereas the outer layer would only shrink $600a$. When the whole block reaches room temperature the inside must, therefore, be in a state of tension and the outside in compression, the strain being roughly proportional to $50a$ per unit length, 50° being the difference in temperature between the outside and the inside of the block at the time when the outside became hard. Such a block would be considerably stronger and the surface harder than if free from strain.

Now let us consider the case of a plate of glass allowed to cool rapidly from the upper surface. A temperature gradient will be set up, the lower surface being, say, 50° hotter than the upper. The upper surface will set hard at, say, 600° and shrink as it sets, and the lower, being still soft, will flow under the compression due to the shrinkage of the upper.

There appears to be a discontinuity in the coefficient of expansion which increases by six or more times its normal value over a small range of temperature corresponding to the "setting" temperature. When the bottom of the block reaches 600° and begins to set hard, the upper surface will have reached 550°. If the temperature gradient has remained constant no strain will be present in the plate. At this point the whole plate has become "set," and any change of temperature gradient will consequently produce strain. As the whole plate cools the temperature gradient will diminish and, finally, the whole block will approximate to room temperature. The lower surface will then be in a state of tension and the upper in compression. If the maximum tension exceeds the tensile strength of the glass, a crack will develop in the lower surface and the glass will break.

(ii.) *Properties of Strained Glass.*—This plate will be strong to an upward force tending to produce concavity in the lower surface and weak to a downward force. Similarly, if scratched with a diamond on the bottom surface it will break easily, but not if scratched on the top. It is, therefore, desirable that glass required for general use should be as free from strain as possible, since the surfaces in tension are very vulnerable. However, for certain purposes some strain would be an advantage; for example, in certain kinds of hollow ware which are not subjected to rough usage on the inside, or in the case of a lamp chimney which has to resist sudden heating from the inside, compression in the outer surface is desirable.

Since glass-ware cooling quickly generally loses heat from the outside, we usually find under-annealed glass with its external surfaces in compression. Thus, when glass tubing, which is seldom annealed after manufacture, is cut into standard lengths, arrangements are made, by means of a diamond suitably mounted on a rod, for scratching the inside of the tube, which then breaks easily along the scratch. This cracking off would present some difficulty in tubing of large bore if it were well annealed; in fact, it would be necessary to heat the glass to produce temporary strain.

(iii.) *Process of Annealing.*—In the process of annealing there are two critical temperatures. The lower critical temperature (600° in the cases cited above) may conveniently be defined as the temperature at which the glass is just deformed at an appreciable rate under a stress equal to its tensile strength when cold. The upper critical temperature may be defined as the temperature at which 95 per cent of the stress in the glass disappears in three minutes.[1] These definitions are, of course, purely arbitrary.

The magnitude of the stresses in the glass will depend on the temperature gradient in the glass as it passes the lower critical temperature.

All glass-ware, with a few exceptions, is annealed after manufacture, generally by passing it through tunnel kilns (lehrs) which are heated at one end. The temperature gradient of these lehrs is so arranged that the glass, when passing through, is heated to

[1] For methods of determining the critical temperatures see *Journ. Soc. Glass Tech.* i. 61 and ii. 90.

its upper critical temperature and cooled slowly down to normal temperatures. For the success of the annealing operation it is, of course, essential that the rate of cooling be slow between the upper and lower critical temperatures, particularly in the neighbourhood of the lower. The rate permissible varies, of course, with the thickness of the ware and the degree of freedom from strain required.

Below the lower critical temperature the rate of cooling is immaterial, provided always that the temperature gradient is not sufficient to cause cracking of the ware.

The annealing temperatures vary largely with the composition of the glass. In the case of lead glasses the upper temperature is about 450° C. and the lower about 300° C.; with soda-lime glasses the upper temperature lies between 500° C. and 600° C. with a lower temperature between 320° C. and 340° C. Some hard resistance glasses require to be annealed at 630° C. to 640° C. and have a lower temperature of about 350° C.

The rate of cooling permissible depends on the thickness of the glass. Near the upper critical temperature the rate of cooling may be fairly rapid, but it must be diminished while passing the lower critical temperature. With a lead glass, a vessel with a thickness of wall, say ¼ in., can be cooled through its critical range at a rate of 50° per hour, whereas an electric-light bulb can be completely annealed and cooled to room temperature in 3 minutes. A soda-lime glass ¼ in. thick can be cooled through its critical range at about 15° per hour, and a complex resistance glass of that thickness would require cooling at less than 10° per hour; in fact, it is extremely difficult to obtain thick samples of the latter glass entirely free from strain, even in the laboratory.

In badly annealed articles it is quite usual to find tensile stresses of 1000 lbs. per sq. in. and often considerably greater.

(iv.) *Detection of Strain.*—Strain may conveniently be detected by means of a polariscope, since glass is doubly refracting when strained. When viewed through a quarter-wave plate between crossed Nicols a slight strain in a glass article may readily be located and the degree of strain estimated.[1]

§ (20) Devitrification.—The term devitrification denotes the formation of an appreciable quantity of a crystalline phase in the glass. The condition which favours devitrification in a glass is supersaturation with respect to one of its constituents or compounds at a temperature at which the viscosity is sufficiently low to permit of rapid crystal growth. The tendency of a glass to devitrify is influenced by factors which are as yet imperfectly understood. There is evidence that gases dissolved in the glass exert a marked influence in this respect; certain glasses, made with raw materials which have been carefully dried before use do not devitrify when worked in the blow-pipe, whereas similar glasses made from a batch not specially dried devitrify readily.

Certain constituents appear to inhibit crystal growth, whereas others appear to act as catalysts in accelerating it. Of the former, alumina is a notable case; boric acid, in certain cases, is also effective. Of the latter class, sulphur trioxide, chlorine, and fluorine are examples.

Of the crystalline phases which are formed, silica, in the form of tridymite, is the most common. In glasses rich in lime, calcium silicate ($CaOSiO_2$) is formed, and, in barium glasses, crystals of barium-disilicate ($BaO2SiO_2$) are deposited. Tridymite generally separates out from lead glasses, but in dense lead flints lead silicate is formed. These crystals may easily be distinguished by their refractive indices:

Tridymite . . .	1·469-1·473
$BaO(SiO_2)_2$ . . .	1·598-1·617
$CaOSiO_2$ . . .	1·621-1·633

Glasses containing soda are, in general, more liable to devitrification than similar glasses containing an equivalent amount of potash.

For the prevention of devitrification in any particular glass the factors controlling the crystal growth must be considered. For example, in the case of a soda-lime glass rich in calcium we may expect the glass to become supersaturated with respect to calcium silicate. The addition of alkali to this glass will increase the devitrification owing to the increased fluidity of the glass at any given temperature, but the addition of silica will increase the viscosity and diminish the devitrification. In the case of a glass rich in silica, say over 73 per cent, the dangerous crystalline phase will be tridymite. The addition of lime to this glass will inhibit devitrification for the reasons given above. The addition of alumina to most glasses prevents devitrification. The formation of crystals is frequently observed on the surface of glasses which show no evidence of crystal growth in the centre of the glass. In some cases this may be due to the volatilisation of alkali from the surface, giving rise to a thin surface film supersaturated with respect to silica. Glasses containing no appreciably volatile constituents show scums, however; for example, a glass containing only CaO, MgO, and SiO_2 gives a scum of crystalline $(MgO)_2SiO$.

Many glasses, when heated in the blowpipe, show a ring, of faintly opalescent appearance

[1] "Annealing of Glass," *Journ. Franklin Inst.* Nos. 5 and 6, cxc.

a short distance from the point where the flame impinges on the glass. This phenomenon is known as "bloom." It has been shown[1] to be due to the reaction of sulphur compounds in the gas with the alkali in the glass. The reaction is at first superficial only and the Na_2SO_4 formed can be wiped off. If the glass is subjected to prolonged heating, a permanent opalescence is produced and can only be removed by melting the glass in the blowpipe, when the product of the reaction is dissolved.

VI. Physical Properties

§ (21).—The measurement of the physical and mechanical properties of glass, as may be expected from the nature of the material, is a matter of some difficulty. The classic researches of Winkelmann and Schott pointed the way to the further study of glass, but it is only in recent years that serious attempts have been made to cover the ground.

In the first place, the properties vary very considerably with the composition of the glass, and it is difficult to obtain specimens with the desired degree of uniformity. Again, the presence of strain exerts a marked influence. We have a further difficulty in that methods of measurement applicable to glass at ordinary temperatures, when for most purposes one may regard it as a solid, have to be abandoned at higher temperatures in favour of those suitable for dealing with liquids. We have therefore a gap in our knowledge of the properties in the neighbourhood of the softening point, where, at any rate from the point of view of the manufacturer and those who have to manipulate glass at high temperatures, exact data is most needed.

In this range there is evidence of a discontinuity in the properties (see "Coefficient of Expansion," § (27)). Heating and cooling curves show that all glasses give an endothermic reaction on heating and an exothermic reaction on cooling. To detect this it is necessary to use a sensitive differential thermocouple, the evolution and absorption of heat being extremely small. This phenomenon occurs at a point when the glass acquires a particular viscosity, and is coincident with the discontinuity in the coefficient of expansion of the glass.[2]

An explanation of these phenomena is suggested by the work of Griffith,[3] which appears to indicate that, near the annealing range, the atoms or groups of atoms which above that temperature are arranged at random rapidly form nuclei possessing a definite orientation. These nuclei or chains of molecules are formed slowly at room temperature.

It is suggested that this arrangement of atoms causes local strain which may produce surface flaws. This theory assumes that the attraction between the molecules is a function of the orientation, since the grouping, to form a crack, must cause a decrease in the potential energy of the system to counterbalance the surface energy of the crack.

Experiments on glass fibres show that a fibre, drawn rapidly from a bead at a high temperature (*i.e.* chilled quickly through the annealing range) has an abnormal strength—10^6 lbs. per sq. in.—of the same order as the "intrinsic pressure," calculated from the energy required for volatilisation.

After a short time, a few hours, this strength decreases, reaching a stable value which depends on the diameter of the fibre. This fall in strength is accompanied by a change in dimensions. With fibres in the stable state, strengths have been obtained from 50,000 lbs per sq. in. in a fibre ·0027″ diameter to 490,000 lbs. per sq. in. in a fibre ·0001″ diameter. By extrapolation, for a fibre of zero diameter—*i.e.* of molecular dimensions—we obtain a figure of about 10^6 lbs. per sq. in.

These strong fibres are only obtained if the glass be drawn hot. If drawn out too cold we have a skin of chilled glass formed surrounding a fluid core; further extension will cause fracture of this skin, giving rise to surface flaws.

The slow decrease of tensile strength at room temperatures, ascribed to molecular rearrangement with the attendant surface flaws, explains the spontaneous cracking of glass articles which sometimes occurs long after they have been made.

Winkelmann and Schott attempted to correlate the physical properties with the chemical composition and, assuming additive relations, to assign factors to the constituents by means of which the physcial constants of any glass could be predicted. The factors obtained were not by any means satisfactory for quantitative work, but, in general, we have been able to discover from them the comparative effects of the constituent oxides on the properties.[4]

§ (22) Density.—The density of glass is approximately an additive function of the composition, and can be expressed in the form of $100/S = \Sigma(p/a)$, where S is the density of the glass, p the percentage of a constituent in the glass, and a is a constant for that constituent. The values of a are not identical with the densities

[1] Travers, *Journ. Soc. Glass Tech.* v. 61.

[2] This affords a convenient method for the determination of the upper critical annealing temperature of an opaque glass. See Scientific Papers of Bureau of Standards, No. 358, "Concerning Annealing and Characteristics of Glass," Tool and Valasek.

[3] "The Phenomena of Rupture and Flow in Solids." *Roy. Soc. Phil. Trans.* A, ccxxi. 163.

[4] Reference should be made to the summary of their work by Hovestadt in his book *Jena Glass*, translated by Everett.

of the oxides; according to Winkelmann and Schott a being always greater. Winkelmann and Schott assigned the following values for a:

B_2O_3	1·9	Al_2O_3	4·1
SiO_2	2·3	As_2O_3	4 1
Na_2O	2·6	ZnO	5·9
K_2O	2·8	BaO	7·0
CaO	3·3	PbO	9·6
MgO	3·8		

The results obtained with these constants give values correct to within about 3 per cent of the observed densities for most glasses. More recent work has shown that the figure assigned to MgO is too high. Assuming a value of 2·3 for SiO_2, Turner obtained a factor of 2·9 for Na_2O and 2·9 for MgO. Tillotson has also calculated factors which give good results: SiO_2 2·3, CaO 4·1, MgO 4·0, Li_2O 3·7, Al_2O_3 2·75.

The presence of strain in glass may cause a variation in the density of one in the second decimal place.

Optical glasses range from 2·3, a light crown, to 5·9, a dense flint.[1]

Common bottle glass	2·46-2·47
Plate glass	2·49
Heavy flint table-ware	3·5

The alkalis and RO group give a low dispersion relative to n_D (and have a high ν), lead oxide gives a high dispersion.

Boric oxide lengthens the red end of the spectrum relative to the blue. Fluorine, sodium, and potassium lengthen the blue end. Barium, while giving a high n_D, gives a low dispersion, and by its use we can obtain glasses with a high ν value relative to their refractive indices. This property is of great service to the optician.

The presence of strain lowers the refractive index. The greater the density of the glass the greater the variation in refractive index due to strain. It is therefore necessary that optical glasses be annealed with the greatest care. In the best glasses, the retardation due to double refraction should not exceed one-sixteenth of a wave-length.

Below are given the optical constants of a few typical optical glasses. The figures given in the last column (taken from Messrs. Chance's list of optical glasses) show the wave-length at which the transmission is reduced to 50 per cent through a plate 1 cm. thick.

Type.	n_D.	n_C-n_F.	ν.	C-D.	D-F.	F-G′.	Density.	Wave-length for 50 per cent abs.
Fluor. crown	1·4785	·00682	70·2	·00202	·00480	·00363	2·47	·301μ
Boro-silicate crown	1·5126	·00818	62·7	·00241	·00577	·00458	2·52	..
Barium crown	1·5881	·00962	61·1	·00284	·00678	·00544	3·31	·358μ
Silicate crown	1·5204	·00869	59·9	·00255	·00614	·00492	2·63	..
Zinc crown	1·5149	·00890	57·9	·00265	·00625	·00506	2·62	·323μ
Baryta flint	1·5515	·01067	51·7	·00310	·00757	·00620	2·99	·323μ
Light flint	1·5632	·01312	42·9	·00375	·00937	·00781	3·07	..
Dense flint	1·6182	·01697	36·4	·00484	·01213	·01031	3·60	·337μ
Heaviest flint	1·9044	·04174	21·7	..	·03023	·02726	5·92	..

§ (23) Optical Properties.—For optical purposes the refractive index for light of wave-lengths ·6563μ, ·5893μ, ·4862μ, and ·4341μ, the C, D, F, G′ lines, is determined, and it is usual to specify the refractive index for the D line n_D and the partial dispersions. The ratio of (n_D-1) to the "mean dispersion" is also generally stated. This quantity $(n_D-1)/(n_C-n_F)$ is known as ν.

For most glasses the refractive index increases with the density and mean dispersion, while the value of ν decreases. Glasses having a high ν are generally designated crown glasses and those with a low ν flint glasses, the dividing line being a ν value of 55.

[1] Winkelmann and Schott, Larsen, *Amer. Journ. Sci.*, 1909, xxviii. 236; Tillotson, *Journ. Ind. Eng. Chem.*, 1912, iv. 246; *Journ. Amer. Cer. Soc.*, 1918, i. 76; Turner and English, *Journ. Soc. Glass Tech.*, iv. 126 and 153.

§ (24) Strength.—The strength of glass varies very considerably with the composition and degree of annealing. The difficulty of handling so brittle a substance when testing has led to very divergent results from different observers.

The tensile strength probably varies between 2½ and 10 tons per sq. in. Under certain conditions glass fibres have been prepared—·00013″ in diameter—having a tensile strength exceeding 214 tons per sq. in.[2]

Crushing strength, 10 to 15 tons per square inch.

Winkelmann and Schott placed the oxides in the following order, as regards their influence on the glass in tension: CaO, ZnO, SiO_2, P_2O_5, B_2O_3, BaO, Al_2O_3, As_2O_5, PbO, Na_2O, K_2O, MgO; in compression, SiO_2, MgO_2, Al_2O_3, B_2O_3, ZnO, PbO, BaO, K_2O, Na_2O, in descend-

[2] Griffith, *Phil. Trans.* A, ccxxi.

ing order of strength. (The positions of CaO and MgO are doubtful.)

§ (25) Hardness.—The hardness of annealed glass on Mohs' scale varies between quartz and fluorspar. Various attempts have been made to assign factors to the constituent by means of which absolute hardness may be calculated, but they have met with little success.

All glasses will scratch any other glass. Auerbach attempted, by measuring the size of the scratches, to arrange glasses in order of their scratching powers and their resistance to scratching. It was found that the order obtained by an estimation of their scratching powers was not by any means identical with their resistance to scratching.

The hardness of a glass surface depends very largely on the heat treatment it has received, surfaces in compression being considerably harder than those in tension (see "Annealing," § (19)). It is claimed that by extremely rapid chilling it is possible to obtain a glass which cannot be scratched by a diamond.

Various processes have been patented for hardening sheet glass by chilling the surface. Rapid chilling can be obtained by pressing the hot glass sheet between copper plates. Glass so treated is said to withstand eight times the shock which would fracture a similar sheet if annealed. The success of such a process depends on the uniformity with which the chilling is effected.

§ (26) Elasticity. (i.) *Young's Modulus.*—Young's Modulus varies from 4·9 to 5·9 dynes per sq. cm. for flint glass and from 5·9 to 7·8 for crown glasses.

Clarke and Turner have calculated factors for various oxides which satisfy the equation E (Young's Modulus) $=\Sigma pz$, p being the percentage of the oxide in the glass and z a constant. This formula gives good results for soda-lime glasses.

Oxide.	Value of z (giving E in kg. per sq. cm.).
SiO_2	40
Na_2O	110
(Al_2O_3)	120
(Fe_2O_3)	120
CaO	240
(MgO)	300

The values for the oxides in brackets require further verification.

(ii.) *Rigidity (Torsion Modulus).*—

Flint glass	$2·0\text{-}2·5\times10^{11}$ (dynes per sq. cm.)
Crown glass	$2·6\text{-}3·2\times10^{11}$ (dynes per sq. cm.)

(iii.) *Volume Elasticity (Bulk Modulus).*—

Flint glass	$3·6\text{-}3·8\times10^{11}$ (dynes per sq. cm.)
Crown glass	$4·0\text{-}5·0\times10^{11}$ (dynes per sq. cm.)

(iv.) *Poisson's Ratio.*—

0·21-0·28.

§ (27) Coefficient of Expansion.—The linear coefficient of expansion varies from $3·3\times10^{-6}$ to 14×10^{-6}, that of fused silica being $0·4\times10^{-6}$.

Typical Glasses.	Linear Coefficient (0°-100°)$\times10^{-6}$.
Heat-resisting glass (cooking ware, etc.)	3·3
Jena 59‴	5·7
Jena 16‴	7·8
English laboratory ware	7·4
Flint glass	7 to 8
Soda-lime glass	7·5 to 9·5 and over.

Turner and English have suggested factors for the calculation of the coefficients of expansion by the formula $a=\Sigma pz\times10^{-7}$, p being the percentage of any constituent and z a constant assigned to it.

Oxide.	Value of Z.
SiO_2	0·05
Na_2O	4·32
CaO	1·63
MgO	0·45

It will be seen that the coefficient of expansion is sensitive to a small change in the alkali content of the glass; an increase of 1 per cent in the alkali may cause an increase of 5 per cent in the expansion.

The expansion is affected considerably by the presence of strain in the glass. Values obtained from strained glass may be 5 per cent higher than those from the same glass after annealing.

The coefficient of expansion rises steadily with the temperature, a linear relation holding over a small range of temperature. With all glasses, however, there is a discontinuity in the expansion near the annealing temperature of the glass, the coefficient increasing by six or even more times its normal value. This abnormality only persists over a short range of temperature, the coefficient of expansion falling at higher temperatures.[1]

§ (28) Specific Heat.—The specific heat of glass at normal temperature may be calculated to within 1 per cent from the specific heats of the constituent oxides, the specific heat $c=1/100\Sigma PC'$, P being the percentage of the constituent and C′ its specific heat. Winkelmann assigns the following values of C′:

SiO_2	·1913	BaO	·06728
B_2O_3	·2272	Na_2O	·2674
ZnO	·1248	K_2O	·1860
PbO	·05118	Li_2O	·5497
MgO	·2439	CaO	·1903
Al_2O_3	·2074	P_2O_5	·1902
As_2O_3	·1276	Mn_2O_3	·1661

The thermal capacity per unit volume shows

[1] "Measurement of Thermal Dilation at High Temperatures," Peters and Cragoe, Bureau of Standards Scientific Papers, No. 393.

only an approximate approach to constancy, varying from about 4·7 to 5·1.

Lead flint, specific heat	·14
Soda lime, specific heat	·19

The specific heat rises with the temperature.[1] Glasses have a specific heat only slightly higher in the amorphous state than in the crystalline (devitrified), with the exception of glasses rich in alkali, where the difference is considerable.

§ (29) Thermal Conductivity. — ·0015-·0025 C.G.S. units. These figures may be taken as approximate for room temperature.

§ (30) Electrical Conductivity and Specific Inductive Capacity.—The conductivity of glass depends largely on the alkali content of the glass. Glasses rich in soda have a high conductivity. The substitution of potash for soda reduces the conductivity.

Ambronn[2] states that the conductivity can be expressed in the form $L = L_0 e^{-\beta/\theta}$, L_0 and β being constants, L the conductivity, θ the absolute temperature. L_0 can be expressed as a linear function of the soda and lime in the glass.

Type of Glass.	Specific Resistance×10[10].	Temperature, °C.
Soda lime .	531·05	55
	89·15	72
	11·901	93
	1·874	116
	0·202	149
	Specific Inductive Cap.	
	6·26	11
	6·79	129
	Specific Resistance.	
Soda lead .	215·13	73
	144·64	83
	20·82	104
	4·93	120
	1·89	140
	Specific Inductive Cap.	
	7·36	19
	8·44	130
	Specific Resistance.	
Potash lead.	1328·6	142
	Specific Inductive Cap.	
	6·76	18

According to Gray and Dobbie,[3] annealing reduces the conductivity of glass by many times its value in the unannealed state.

§ (31) Thermal Endurance. — Winkelmann and Schott adopted the formula

$$F = \frac{P}{aE}\sqrt{\frac{K}{SC}},$$

F being the resistance to a sudden change of temperature,
P = tensile strength,
a = coefficient of expansion,
E = Young's Modulus,
K = thermal conductivity,
S = density,
C = specific heat.

Values calculated by this formula are more or less in accordance with experimental results.

Since the coefficient of expansion changes more rapidly with composition than do the other physical constants, the thermal endurance is therefore largely governed by the value of a. Glasses rich in alkali are therefore the most liable to break with changes of temperature. Glass being considerably stronger in compression than in tension, it follows that it will be more resistant to sudden uniform heating than to sudden cooling. Glasses which will not withstand a rapid fall of 50° without cracking can be heated rapidly to several hundreds of degrees without risk. It is probable that the condition of the surface of the glass plays an important part in the thermal endurance.

Tests of the thermal endurance of hollow glass-ware may be carried out as follows: A beaker or flask of the glass under test is filled with wax and heated to a degree or so above the temperature of testing. It is then allowed to cool, keeping the wax well stirred, until the required temperature is reached, when it is plunged into iced water. This operation is repeated at successively higher temperatures until the beaker cracks. This test can usually be made to repeat to within 15° C.

Good beakers should stand 150° C. without cracking under this test. Laboratory ware containing striae are found not to be inferior to those free from cords, although the former must be in a state of considerable strain.

§ (32) Surface Tension.—The surface tension of glass[4] has been measured by Griffith at temperatures up to 1100° C. and a value at 15° C. obtained by extrapolation.[5]

Temperature, °C.	Surface Tension in lb. per Inch.
1100	·00230
905	·00239
801	·00257
745	·00251
15 (extrapolated) . .	·0031

[1] White, *Amer. Journ. Sci.* xxviii. 334.
[2] Ambronn, *Phys. Zeitsch.*, 1913–14, 112; 1918–19, 401.
[3] Gray and Dobbie, *Proc. Roy. Soc.*, 1898, lxiii. 38; 1900, lxvii. 197.
[4] The glass used had the following composition: SiO_2, 69·2 per cent; K_2O, 12 per cent; Na_2O, 0·9 per cent; Al_2O_3, 11·8 per cent; CaO, 4·5 per cent; MnO, 0·9 per cent.
[5] Griffith, *Phil. Trans. Roy. Soc.* A, ccxxi.

§ (33) Viscosity. — The viscosity of glass falls rapidly with a rise in temperature. There are great experimental difficulties associated with the measurement of viscosity at high temperatures, and there is but little data on the absolute viscosity of glass at present available. Practical glass-makers have, however, collected information as to the general effect of certain constituents on the melting temperature and rate of setting, *i.e.* the temperature coefficient of viscosity. Thus, the addition of silica decreases the temperature coefficient, *i.e.* glasses rich in silica have a long working range, whereas the addition of lime increases the temperature coefficient. Alumina and baryta are also said to impart a high temperature coefficient to glass.

The viscosity of glass at the "fining" temperature is of the order of 50 (absolute units). At the gathering temperature the viscosity is from 1500 to 3500, depending on the size of the ware manufactured and the method of blowing. With automatic blowing devices the glass is gathered at a much lower viscosity. The viscosity at the annealing temperature is 10^{13}-10^{14}.

The measurement of viscosity has been used for the determination of the annealing temperature, the method adopted being the observation of the rate of twisting of a glass rod under a known torque.[1]

VII. Coloured and Opal Glasses

§ (34).—Coloured glasses may be divided into two groups, according as the colouring agent is (i.) apparently in true solution in the glass or (ii.) suspended in a fine state of division resembling a colloidal solution. As in the case of aqueous solutions, there is no sharp line of demarcation between the two groups.

Coloured glasses in the category (i.) are produced by the oxides of iron, titanium, manganese, cobalt, nickel, chromium, copper, etc., and in the second group (ii.) by the elements carbon, copper, gold, and selenium, etc. The opal glasses may be considered as coming under the heading (ii.).

(i.).—The colours produced by the various oxides vary somewhat with the composition of the glass and the conditions of melting (*i.e.* oxidising or reducing). In the case of the green glasses, the soda glasses tend to a bluish-green and the potash to a yellowish-green.

(*a*) *Iron.*—In the divalent state iron imparts a bluish-green colour, in the trivalent condition a greenish-yellow is produced. The colour from the divalent iron is much more intense than that from the trivalent, and hence oxidising conditions of melting are conducive to freedom from colour.

(*b*) *Titanium* oxide in combination with iron is used for the production of amber glass; very beautiful shades of amber may be obtained.

(*c*) *Manganese.* — Manganous compounds have very little colouring action on glass, giving a faintly green colour in small concentrations and a brown colour in large. When oxidised, colours ranging from pink to purple are produced. In small quantities manganic compounds neutralise the blue-green colour due to iron and are, accordingly, frequently used for this purpose, where oxidising conditions prevail. Used in larger quantities with iron, amber glasses are produced. Manganese is used extensively in combination with other oxides where a purple tint is required.

(*d*) *Nickel.*—The colouring effects of nickel are very variable; in fact, it has been claimed that with this oxide alone all the colours in the spectrum can be produced, with suitable modifications of the batch mixture and firing conditions. Nickel is little used, therefore, commercially, although it is sometimes used with considerable success as a decoloriser to neutralise the effects of ferric iron.

(*e*) *Cobalt* yields intense blue glasses. 0·01 per cent of the oxide is sufficient to produce a pale blue and 0·1 per cent a medium blue glass. The action of cobalt is similar in reducing and in oxidising conditions, and it is used almost exclusively for the production of blue glasses. Cobalt glasses, however, transmit an appreciable proportion of red rays. The addition of a small quantity of chromium cuts out the red rays (1 part of Cr to 10 Co).

(*f*) *Chromium* produces intense green glasses, but the colouring effect is not so strong as in the case of cobalt. It is used extensively in the industry, since the chromium colours are little affected by melting conditions. Chromium is generally introduced in the form of potassium bichromate.

(*g*) *Copper.* — Divalent copper gives tints varying between blue and green. Under reducing conditions ruby glass can be obtained (see below).

(*h*) *Iridium* imparts a neutral grey tint. In large quantities a black glass can be obtained.

(*i*) *Cadmium*, in the form of the sulphide, produces a rich yellow glass.

(*j*) *Uranium* also gives a yellow glass, characterised by a marked fluorescence in any but lead glasses.

[1] Littleton and Roberts, *Journ. Opt. Soc. of Amer.* iv. 4; Trouton and Andrews, *Proc. Phys. Soc. Lond.* xix. 47.

For data on the viscosity of glass and slags see Arndt, *Zeit. für Electrochemie*, 1907, p. 572; *Chem. Apparatenkunde*, 1903, No. 3; Field, *Trans. Far. Soc.*, 1917–1918, xiii.; Washburn, *Phys. Rev.*, Feb. 1920, xv. 2.

(*k*) *Antimony* may produce yellow glasses when used with lead batches.

(ii.).—In the second group of colouring agents copper, gold, and selenium are used commercially for the production of ruby glass. The manufacturing process presents some technical difficulty.

(*a*) The copper ruby is the cheapest to make, but the process requires great care, and frequently inferior colours are produced. It is necessary that the copper be reduced to the metallic state, and the presence of a reducing agent is necessary. Partial oxidation imparts a muddy green colour to the glass. The copper is usually introduced into the batch in the form of the oxide. When the glass is fine the copper is apparently in complete solution, and proofs taken from the pot and chilled quickly may be colourless. The glass is then gathered and blown to the required shape. The article is then reheated and the ruby colour comes up (the glass being said to "strike"). With the copper ruby the colour is so intense that it is necessary to use the ruby glass in thin sheets only, the articles required being "flashed" with ruby (*i.e.* made from colourless glass and covered with a thin coating of ruby; this is effected by gathering the glass for blowing the article from a pot containing "white" glass and then, before blowing, dipping the gather in the ruby glass; by this means glass-ware can be made with an extremely thin coating of ruby glass). When the glass is reheated the particles begin to aggregate, and when the glass "strikes" they have attained a sufficient size to produce scattering of the light. If the ruby glass is heated further the colour will change to purple, and from purple to blue as the particles increase in size. Finally, if heated for a sufficient time at a suitable temperature, the particles become visible and the well-known copper aventurine glass is produced. To obtain a good ruby the degree of dispersion of the metallic copper lies within fairly narrow limits.

(*b*) The gold ruby is made in the same manner, but the glass is much easier to handle, since there is not the same tendency to oxidation as in the case of the copper ruby, and it is possible to work with small concentrations of gold. It is therefore not necessary to "flash" gold ruby. The gold is usually introduced in the form of gold chloride.

The presence of tin oxide greatly facilitates the striking of the ruby and is almost invariably used in the industry with the copper ruby, and generally with the gold ruby.

(*c*) Selenium in the metallic state is also used for the production of ruby glass. In small quantities selenium is used as a "decoloriser" under conditions sufficiently reducing to ensure that the selenium remains in the metallic state. It is used for this purpose in tank furnaces. The oxide of selenium exerts no colouring action.

(*d*) Carbon, in suspension in glass, gives colours ranging from yellow to amber.

(*e*) Many special glasses have been made for the transmission and absorption of light of particular wave-length. Thus there are at present on the market a large variety of glasses similar to Crookes' glasses, the principal constituent of which is cerium oxide, which transmit nearly all the light in the visible spectrum, but are particularly opaque to ultra-violet light and cut down the infra-red light considerably. The absorption of these glasses for ultra-violet light is said to be improved by exposure to a strong source of ultra-violet. Similarly, nickel glasses are made which are opaque to visible radiation but are highly transparent to ultra-violet light.

(*f*) The colour of glass is modified by the prolonged action of sunlight and ultra-violet light. Glasses containing manganese rapidly develop a purple colour on exposure to intense ultra-violet light. This colour disappears if the glass is heated to its softening point, but reappears again if exposed further to ultra-violet light. Soda-lime glass tends to develop a blue colour, nickel and selenium a yellow.

(*g*) A rise in temperature displaces the absorption in coloured glasses towards the red end of the spectrum.

§ (35) OPAL GLASSES.—Opal glasses are made under conditions somewhat similar to the manufacture of ruby glass, the opalescence being due to the presence of small particles in suspension. Some of the opals, however, are regarded as being emulsoids rather than suspensoids, the disperse phase being a glass insoluble at low temperatures.

(i.) *Fluoride Opals.*—The most common opacifying agent is fluorine. This may be introduced as fluorspar (CaF_2), sodium fluoride (NaF), or cryolite ($3NaF, AlF_3$). In practice the presence of alumina serves to facilitate the production of the fluoride opals, although these can be made without alumina. The most usual ingredients in an opal batch are a mixture of fluorspar and felspar or cryolite alone, or a mixture of the three. Up to 20 per cent of cryolite may be used in an opal batch. The character of the opal varies somewhat with the form in which the fluorine is introduced. Thus, the NaF opal is faintly opalescent over a long range, whereas the cryolite opal strikes rapidly. By a careful control of the quantity of fluorine present opals can be made to remain clear when cooled quickly and to strike when reheated. Various ornamental effects are thus introduced into artistic ware by judicious local

heating. The composition of the suspended particles is not yet known with certainty.

(ii.) *Phosphates.*—Good opal glasses are also made with calcium phosphate (bone ash) as the opacifying agent. These opals strike at a higher temperature than the fluoride opals; they require a higher temperature to melt, and are not so easy to manipulate. They have the advantage, however, that they do not tend to "boil" readily if overheated, as do the fluoride opals, which evolve fluorine copiously if the temperature be allowed to exceed 1300° C.

(iii.).—There are other materials sometimes used in the manufacture of opals. Arsenic yields a good opal, and oxide of tin is used, especially in the enamelling industry.

The opal glasses are usually fragile and break easily if exposed to a rapid change of temperature. If reheated for any length of time, or worked in the blow-pipe, the suspended particles grow, especially if the concentration of the opacifying agent in the glass is high, and the glass has the appearance of devitrification, the surface losing its gloss. If examined microscopically it is seen that the particles consist of globules of a glass having, as can be determined by manipulation in the blow-pipe, approximately the same softening point as the matrix.

Opalescent effects are often found as the result of incipient devitrification, the crystallisation of silica sometimes producing this effect.

References

H. Siedentopf and R. Zsigmondy, "The Observation and the Determination of the Magnitude of Ultra-microscopic Particles, with particular Application to Gold Ruby Glass," *Ann. Physik*, 1903, x. 1.

J. C. M. Garnett, "Colours in Metal Glasses and in Metallic Films," *Phil. Trans.*, 1904, A, cciii. 385; 1906, A, ccv. 237.

M. Luckiesh, "The Physical Basis of Colour-Technology," *Journ. Franklin Inst.*, July 1, 1917.

G. Chesnau, "Contribution to the Chemical Study of Coloured Glass of the Middle Ages," *Compt. Rend.*, 1915, clx. 622.

C. Dralle, "The Colour which Iron and Manganese Compounds impart to Glass," *Sprechsaal*, 1901, xxxiv. 68, 103; *Chem. Zeitung*, 1900, xxiv. 1132.

R. Zsigmondy, "On the Absorption of Light in Coloured Glasses," *Ann. Physik*, 1901, iv. 60.

R. Zsigmondy, "Coloured Glasses for Scientific and Technical Purposes," *Zeitsch. Instrumentenkunde*, 1901, xxi. 97.

J. W. Mellor, "A Note on Blue Glass," *Trans. Eng. Cer. Soc.*, 1918–19, ii. 478.

W. D. Bancroft, "The Colours of Colloids, Glasses, Glazes, and Gems," *Journ. Physical Chem.*, 1919, xxiii. 603, 623.

VIII. Miscellaneous and Ornamental Glasses

§ (36) Cut Glass.—Cut glass is in common use for many purposes—table-ware, massive glass-ware, etc. This industry is larger than is generally supposed. The preparation of cut glass-ware is a laborious operation requiring considerable skill. The first step is the preparation of the "blank," which may either be pressed or blown, generally the latter. The pattern to be cut is then painted on the blank; these patterns are often very complicated. Sometimes a transfer is used for this purpose. The ware is then taken to the grinding shop and ground on a coarse stone or an iron wheel fed with sand. After the coarse grinding the glass is finished on a finer grade of emery or carborundum wheel, and finally polished on a wooden wheel fed with putty powder (tin oxide). In the case of heavy articles this work is very laborious, a heavily cut piece taking some weeks to complete.

After the fine grinding the polishing is sometimes effected by "acid polishing," thereby avoiding the use of the wooden wheel and putty powder. In this process the ware is dipped into a solution containing hydrofluoric acid. If the composition of the solution is properly controlled a beautiful finish is produced on the glass—with a great saving of time and labour.

§ (37) Engraving.—This form of decoration is frequently applied to "flashed" glass, but is also employed for colourless glass. The engraving is carried out by means of a small soft iron or copper disc, usually revolving at a high speed, fed with fine emery or carborundum powder. In the hands of a highly skilled engraver very beautiful and delicate effects may be produced.

§ (38) Etching and Embossing.—Most glass-ware may be etched by means of a solution of hydrofluoric acid or sodium fluoride. A weak solution of hydrofluoric acid leaves a polished surface; strong hydrofluoric and sodium fluoride give matt surfaces, the latter reagent yielding a finer grained surface than the former.

In the etching process the glass is first coated with Brunswick black or wax, leaving the parts to be etched uncovered. This is effected either by means of a transfer or by coating the whole surface with wax and cutting away the part to be etched. For small designs a pantograph is sometimes used for this purpose. The glass is then immersed in the etching solution and afterwards washed. The wax may then be washed off with hot soap and water, the Brunswick black being removed by washing with turpentine or a solution of caustic soda.

The solvent action of a weak solution of hydrofluoric acid affords an excellent method for cleaning glass surfaces. Glass rinsed with a weak solution is cleaned almost instantaneously without an appreciable loss in weight.

§ (39) Sand-blasting.—This is the cheapest method for preparing a ground-glass pattern on glass surfaces. It is commonly used for labelling bottles, etc. A stream of sharp-

grained sand is blown against the glass by means of a steam- or air-blast, and a ground surface is rapidly produced on those parts not protected. The glass to be sand-blasted is covered with a stencil plate cut, from metal foil or parchment, to the desired pattern. For elaborate patterns dextrine, applied by means of a stencil, is sometimes used as a protection.

§ (40) Frosted Glass. — So-called frosted glass is made by coating the surface of glass, which has previously been sand-blasted, with glue, which, on drying, contracts and chips off flakes of glass, producing the appearance of a frosted window-pane.

§ (41) Beads.—Small glass beads are usually made from thick-walled tubing. Short lengths are cut from the tubing and these are heated in a revolving drum until the edges are rounded off. To prevent the beads from adhering to each other and the holes through them closing up, they are kneaded with clay or talc before heating. They are subsequently polished by being shaken with bran. Glass marbles are made in a similar way, using rod instead of tubing.

Opaque glass for making beads is sometimes obtained by stirring the glass rapidly before drawing the tube. By this means a large number of small bubbles are introduced into the glass, which somewhat resembles mother-of-pearl.

Larger beads and imitation pearls are made by blowing small bulbs from tubing. Coloured pigments are then introduced inside the bulbs. In the manufacture of imitation pearls "essence of pearl" is used. This is made from scales of fish ground to an impalpable powder. The silver and coloured balls (associated with Christmas trees) are made by silvering the inside of the bulbs and then painting the outside with a transparent enamel.

§ (42) Drilling and Slicing.—Glass may be drilled by means of a sand-blast. For finer work, a hard steel drill, lubricated with turpentine, or a copper or brass tube, fed with carborundum or emery, may be used.

Blocks of glass may be cut rapidly by means of a diamond circular saw—a soft iron or mild steel disc armed with diamond dust.

§ (43) Reinforced Glass. — Various processes have been employed for strengthening sheet glass. Of these, the most common is the use of wire-netting embedded in the glass. The wire, which is pre-heated, is rolled into the glass during the manufacture of the sheet.

Another process now in operation consists in cementing together thin glass sheets by means of a thin layer of celluloid. A composite sheet made in this way will not splinter when broken.

§ (44) Fused Silica.—For the properties and uses of fused silica, or quartz glass, see article on "Refractories." E. A. C.-P.

Bibliography

The bibliography of glass is scattered widely throughout scientific literature. References have been given, under separate heading above, to the most reliable sources of information. Below is a list of the principal works on glass:

Powell and Chance, *Principles of Glass-making.*
Rosenhain, *Glass Manufacture.*
"Glass," articles in *Encyclopædia Britannica*, 11th ed.
Hovestadt (translated by Everett), *Jena Glass.*
Shenstone, *Methods of Glass Blowing.*
Journal of Society of Glass Technology. (This Journal contains comprehensive abstracts from current literature.)
Transactions of the American Ceramic Society.
Henrivaux, *Le Verre et le crystal.*
Henrivaux, *La Verrerie au XXieme siècle.*
Gerner, *Die Glasfabrikation.*
Dralle, *Glasfabriken und Hohlglasfabrikation.*

Glass, Action of Steam on. See "Glass, Chemical Decomposition of," § (2).

GLASS, CHEMICAL DECOMPOSITION OF

§ (1) Devitrification and Colour Changes. (i.) *Atmospheric Agents.*—All glass is to a greater or less degree subject to change after manufacture. These changes may arise from chemical transformations occurring in the body of the glass itself, or may be produced by agencies outside giving rise to modifications of the surface of the glass; in both cases the glass is considerably altered, and where the transparency is of prime importance the glass may be rendered quite unserviceable.

The chief change arising in the body of the glass itself is known as devitrification, and is caused by the crystallisation of various metallic silicates, or more rarely silica, giving rise to definite crystalline aggregates, or in extreme cases a micro-crystalline structure affecting the whole mass of glass and rendering it opaque. These phenomena are usually noticed in batches of glass which have been allowed to remain for prolonged periods at too low a temperature during melting, or on reheating for the purposes of working or annealing; during any of which operations the range of temperature for the formation of the various silicates has been maintained for prolonged periods, thus favouring their crystallisation and separation. Recent researches made on the cooling of molten silicate mixtures and the thermal effects observed during the process have enabled some of these critical ranges to be avoided, with the result that devitrification has become a less serious problem. Various impurities appear to induce devitrification, while the favourable influence of others in preventing the trouble (*e.g.* alumina) have been known for a long time.

Glass also undergoes certain chemical changes from within due to other causes than those of devitrification.

In the manufacture of glass of the highest quality for optical purposes, or for use where the maximum transparency and lack of colour is of prime importance, the greatest care has to be taken to ensure that all the materials used are of the highest degree of purity. For example, it has been found that the presence of chlorides or sulphate in the alkali used in the manufacture of potash lead glass is capable of producing cloudy glass unless certain precautions are observed during the process of manufacture. By employing a high temperature, or by the addition of borax to the melt, the trouble may be obviated. It is, however, not impossible to obtain alkali which is practically free from these impurities, and this course is to be preferred.

Another impurity which is impossible to remove entirely, and which when present in any but the most minute proportions (0·03 per cent) produces a distinct coloration, is oxide of iron. The coloration can be destroyed by the addition of peroxide of manganese. The addition, however, frequently causes a secondary change to supervene, giving rise to the formation of a pink colour. Several explanations have been given of the phenomenon. The first action of the oxide of manganese is to oxidise the iron, producing a nearly colourless ferric silicate, which under the influence of sunlight, or the ultra-violet rays, is capable of again becoming deoxidised, giving rise to a higher and coloured oxide of manganese. It has also been supposed that the colour is due to the formation of alkaline permanganates. The strong pink colour arising from this change is commonly noticed in old window glass. In the manufacture of the best optical glasses little or no manganese is added, but every precaution is taken to ensure that all the materials are as free from iron as possible. Glass which has become coloured by manganese in this way is rendered colourless on again remelting. The presence of some base which is capable of forming a higher and lower oxide seems necessary before the influence of manganese makes itself felt.

(ii.) *Action of Water*.—In addition to the changes arising in the mass of the glass itself, there are others which are caused by outside agencies and affect its surface, resulting in the production of tarnish, or actual disintegration of the glass material. The most important substance which is capable of producing these changes in glass is undoubtedly water, and in some cases its action has been so pronounced that otherwise useful optical glasses have had to be discarded, or to be enclosed between layers of more durable glass to make them suitable for employing in optical instruments. In the case of glass used for optical purposes it is found that moisture is actually absorbed from the atmosphere, due to the hygroscopic nature of the glass; the water is capable of causing solution of the glass with a simultaneous liberation of alkali, which on combination with the carbon dioxide of the air produces a layer of alkaline carbonates. It has been observed[1] that glasses containing potash, lime, and silica are very much more hygroscopic, and therefore more susceptible to attack by this means, than are the corresponding soda glasses. The presence of increasing amounts of lime decreases this disparity to a great degree. Glasses containing oxide of lead are on the other hand far less hygroscopic than either, and are therefore not so liable to decomposition by this means.

The changes produced by water on glass are more pronounced as the opportunity for contact is greater, and in consequence glass exposed to water is attacked more rapidly than when moist air only is in question. Rise of temperature, particularly if accompanied by rise of pressure, as in the case of water-gauge glasses, considerably increases the rate of the attack. Glasses containing boric acid, alumina, or zinc, with soda as alkali, have proved to be particularly resistant to the attack by water at all temperatures, and are, moreover, capable of withstanding prolonged contact with water at high temperatures and pressures, without undergoing marked deterioration, or becoming obscured by the liberation of opaque insoluble basic silicates, as in the case of the lime alkali glasses. The two well-known Jena glasses 16‴ and 59‴ are of this type. Another change affecting optical glass in which water undoubtedly plays a part, but at the same time is not the sole cause of the phenomenon, is what is generally alluded to as "filming." It is found that the interior lenses, prisms, and other glass parts of optical instruments occasionally become veiled, causing a diminution of the light transmitted, and in extreme cases almost complete opacity. The exact cause, or causes, has not been clearly elucidated, but whatever they may be, there appears little doubt that the glass has suffered some chemical decomposition.

When "filmed" glass is examined under low magnifications, at least two distinct types of films are noticed; in the one, distinct globules may be distinguished, while in the other the glass is obscured by a bloom which under higher magnifications is also seen to be globular. The "film" is frequently noticed to be grouped round minute scratches or unevennesses; on the other hand it may very

[1] Foerster, *Berichte der Deutsch. Chem. Ges.*, 1893, xxvi. 2920.

markedly avoid them. On exposure to air the film does not spontaneously evaporate, but does so on heating, leaving a residue and usually a tarnish of the glass surface, with or without actual pitting. It has been proved that the lubricants used in the lens threads, and the varnishes or blacking compounds used on the interiors of the cells or chambers, play some part in the process, and that such compounds which give off volatile constituents at ordinary or slightly elevated temperatures are unsuitable for use in optical instruments. Absolute cleanliness in the polishing and subsequent treatment of the lens, together with drying of the air in the chambers before sealing up, and the use of suitable lubricants, appear to prevent filming to a great degree. The treatment of the lens with hot water after polishing is also claimed to act as a palliative, but with all these precautions the prevention of filming cannot yet be said to be certain, and more research is required. One thing, however, seems conclusively proved, namely, that water plays a large part in the process.[1] Atmospheric dust is also capable of causing the decomposition of the polished surfaces of glass lenses. The degree of attack depends in the case of glasses free from lead on the amount of alkali present. When this is below 10 per cent it is found on microscopic examination that each dust particle has become a centre of decomposition, while in the case of glasses richer in alkali the whole surface may undergo a homogeneous decomposition with the liberation of alkaline salts, forming deliquescent globules in the case of potash glasses, and a dry dusty deposit in the case of soda glass. In glasses containing more than 20 per cent alkali the deposit becomes visible to the naked eye.

Lead glasses containing more than 20 per cent oxide of lead exhibit what are known as lead spots; these on microscopic examination are seen to originate from dust particles, each of which forms the centre of a mass of minute brown or blackish scales. They are presumably lead mirrors produced by the reducing action of the dust on the solution of the glass formed by traces of moisture.[2]

§ (2) DECOMPOSITION CAUSED BY CHEMICAL REAGENTS.—It will be seen from what has been said above of the various factors affecting the stability of glass, that it becomes necessary to subject that intended for optical or other scientific purposes to test before use in order to prove whether it is suitable for the purposes for which it is intended.

The optical glasses offer particular difficulties, since it is by the variation of the chemical constituents that the special optical features are obtained; consequently the determination of what changes are permissible without seriously affecting the chemical stability becomes a matter of prime importance. Where the optical constants or slight colour can be ignored there is more scope for variation, and it becomes possible to make glasses which are more resistant to atmospheric agents, and at the same time to manufacture those possessing in a high degree resistance to the more violent attack of chemical reagents.

Before passing on to the testing of glassware, it will be advisable to consider briefly the influence of the various constituents of the glass on the resistance to attack by chemical reagents.

(i.) *Action of Water and Steam.*—In addition to ordinary atmospheric moisture, glass for many purposes has to withstand the action of hot and cold water, and also steam, sometimes under high pressures.

The attack proceeds in a similar manner as in the case of atmospheric moisture, but at a greater rapidity. Glasses containing boric acid with either zinc or alumina, and soda as the alkali, have, as mentioned above, proved to be by far the best glasses. When glasses containing these elements are examined it is found that some which may be superior at low temperatures (20° C.) may not be so good as others (at, say, 80° C.), while at still higher temperatures the superiority may be assumed by another glass. For ordinary chemical ware the most useful proportion of boric acid appears to be 9-10 molecules to 100 SiO_2, while for gauge glasses a slightly greater proportion is advisable. The lime alkali glasses are very much inferior, the degree of attack depending largely on the ratio of lime to alkali; the richer the glass is in alkali the less the degree of resistance. Glasses containing more than 20 per cent alkali cannot be regarded as suitable for use with water under any circumstances. The proportion of lime has a marked effect in increasing the resistant powers, and more particularly if the soda is partly replaced by potash, as in the glass used by Stas in his classic researches, in which case the glass compares favourably with the boric glasses.

Lead glasses are intermediate in position between the boro-silicate and lime alkali glasses, the resistant power increasing with the lead content.

(ii.) *Action of Acids.*—Sulphuric, nitric, and hydrochloric acids when concentrated have very little action on most good glasses of the lime soda or lead types, but when the acids are weaker the action becomes slightly more pronounced, and increases on further dilution. This observation has led to the opinion that

[1] Ryland, *Trans. Optical Society*, 1918, xix. 178; Martin and Griffiths, *Trans. Optical Society*, 1919, xx. 135.

[2] Zschimmer, *Chemiker Zeitung*, 1901, xxv. 69.

the attack proceeds from the contained water rather than from the acid itself, and is less than in the case of pure water under similar circumstances, owing chiefly to the neutralisation of the liberated alkali by the acid. The alkali would normally in the case of the action by water be instrumental in further increasing the attack.[1] The boric acid glasses are not so resistant to hydrochloric as to sulphuric acid, and more recent work than Foerster's has tended to emphasise the importance and significance of the attack by acids, more particularly hydrochloric. It has been shown that fuming hydrochloric acid has a very pronounced action on zinc and lime borosilicates, and that high boric acid content tends to promote this attack.[2] On the other hand, glasses containing lime and alumina with soda, and some potash, are much more resistant. A high silica content (70 - 72 per cent) appears to confer great resistant properties.

The conclusion arrived at by Foerster, that in the case of dilute acids the action appears to be due to the contained water, rather than to the acid itself, is confirmed by recent work. The reason why hydrochloric acid should be so much more active than other acids is somewhat obscure.

Lead crystal glass (33 per cent oxide of lead) is very slightly attacked by acids and its resistant power is increased on prolonged contact. The very rich lead glasses are rather severely attacked. The rate of increase of attack with rise of temperature is not so pronounced in the case of acids as with water.

(iii.) *Action of Caustic Alkalies.*—The action of water on glass results in the production of an alkaline solution; we should therefore expect that there is no clear line of demarcation between the results obtained with water and solutions of the caustic alkalies. In the case of dilute solutions this is the case, and the boro-silicates again occupy the premier position as regards resisting power. If, however, the boric acid is greater than 5-7 per cent the resistance diminishes and becomes more pronounced with increasing concentration of alkali.

The lime alkali glasses are attacked by dilute alkalies with the formation of basic lime silicates which appear as opaque layers on the glass, while the solution is found to contain silica and some lime. When the concentration of the alkali increases and exceeds 2N the glass is dissolved as a whole and very little tarnish occurs. The effect of rise in temperature is very marked in the case of the caustic alkalies, more particularly above 80° C. Soda is the most violent in its attack of all the caustic alkalies. Ammonia also has a slight action on glass, but its action has not been so systematically studied; in general, however, glasses which are most resistant to attack by caustic soda are also found to exhibit the same order when ammonia is substituted.

(iv.) *Action of Alkaline Carbonates.*—The attack by alkaline carbonates proceeds in a similar manner as with the caustic alkalies, but in the case of the lime alkali glasses, varieties containing alumina are far more resistant than those which do not contain it. In fact, the alumina-free lime alkali glasses are more strongly attacked by sodium carbonate than by sodium hydroxide of equivalent strength, which is not the case with any others. In explanation of this it has been suggested that as alumina is soluble in caustic soda, but not in sodium carbonate, this is sufficient to account for the difference in behaviour. There appear to be other factors which modify the results, the nature of which is not exactly known.

Solutions of salts also have a slight solvent action on glass, but the degree of attack is in no case comparable to that observed with the reagents considered above.

When the chief factors affecting the stability of glass which is required to withstand the attack of chemical reagents are considered, it will be found that the introduction of one element may, while conferring stability as regards one mode of attack, be conducive to more accelerated attack by a different reagent It therefore becomes a matter of some difficulty to indicate the ideal glass for all purposes, but nevertheless, as a result of exhaustive tests, certain definite conclusions have been reached. Boric acid, for example, besides conferring on a glass great powers of resistance to the influence of water, at the same time confers very valuable mechanical properties, such as power to withstand sudden changes of temperature, which is of great service in all glass articles. It will therefore be found that, in spite of certain minor disadvantages, nearly all glass required to be used for chemical purposes contains a certain proportion of boric acid. The influence of a high silica percentage in conferring acid-resisting powers, and of alumina in restraining the attack by sodium carbonate, have also been alluded to. It will thus be seen that sufficient information is available for selecting the most suitable type of glass for use with any *single* chemical reagent, and considerable progress has been made in the manufacture of glasses which successfully withstand the attack by most chemical reagents to a very high degree.

§ (3) Testing of Glass. (i.) *Optical Glass.*—Glass intended for optical purposes must, in

[1] Foerster, *Zeitsch. anal. Chem.*, 1894, xxxiii. 299.
[2] Cauwood English and Turner, *Journal Society of Glass Technology*, 1917, p. 189.

addition to possessing the properties necessary for the construction of the best optical instruments, be capable of retaining its polish and transparency under ordinary conditions of use. As has been pointed out above, the chief agent in promoting the changes in optical glass is atmospheric moisture, and all tests which are in use for testing optical glasses have been devised to measure directly, or indirectly, the degree of attack.

In the early experiments powdered glass was heated in water and the amount of alkali and total solid material dissolved was determined. This method is not an ideal one, owing to the fact that it is impossible to ensure uniformity in the degree of fineness of the various samples of glass. As a result of these experiments, however, it was found that the less stable glasses gave up a greater amount of alkali and total solid matter, and that the ratio of alkali to silica in the dissolved glass material was invariably higher than in the original glass. Mylius and Foerster,[1] who were the first to investigate this question systematically, were led to conclude that a determination of the alkali removed would give a measure of the relative stability of glasses. They were therefore led to investigate methods for the detection of minute quantities of alkali. As a result they developed the use of a reagent which is now invariably employed for the purpose of classifying optical glasses.

Iodeosin (tetra iodo fluorescein), the reagent in question, is a red solid which is soluble in aqueous ether, but insoluble in water. The aqueous ether solution when shaken up with a solution of an alkali produces a solution of the salt of iodeosin which is soluble in water, giving rise to an intense red colour. The alkali may be determined either by comparing the solution with one containing a definite quantity of the alkaline salt of iodeosin, or by titration with N/100 or N/1000 acid. In the latter case the acid is added to the solution, which must also contain an excess of an aqueous ether solution of reagent, and the two layers shaken together until the water solution is rendered colourless. The extreme sensitiveness of this reaction has enabled the test to be applied to quite small areas of glass, and there is no longer any necessity to use powdered glass for the purpose.

The method as finally developed for the testing of optical glass will now be briefly described, but for a more detailed description the original[2] papers should be consulted.

For the purpose of the test, polished samples of optical glass measuring $6 \times 10 \times 0{\cdot}8$ cm. are used. These are very carefully cleaned and then broken into two portions by making a fine file mark parallel to the longest side and placing a red-hot rod along the cut, followed by a damp piece of filter paper. Any glass dust is removed from the fractured surface by means of a clean brush, and the specimens are now placed in a closed vessel containing water and arranged that the fractured surfaces of the glass are supported in a horizontal position on a platform above the water. The glass vessel and its contents is now maintained at a constant temperature of 18° C. for one week. The glass surfaces are attacked by the moist air to a greater or lesser degree with liberation of alkali, and in the case of very poor glasses it is sometimes found that the surface becomes covered with visible spots of alkaline salts. The specimens of glass are now removed with a clean pair of forceps, and held broken edge downwards in a bath of the iodeosin reagent for one minute. The iodeosin reagent is usually prepared from the sodium salt, in which form it occurs in commerce. It is sometimes found to be adulterated with other salts and dyes, and care should be taken that only the purest samples are used. A weighed quantity of the salt (0·53 gm.) is dissolved in 30 c.c. distilled water and treated with 15 c.c. $(N/I)H_2SO_4$ in a separating funnel. The acid liberates the free dye in the form of orange-red scales. One litre of aqueous ether of the highest purity saturated with water at 18° C. is now added, and the lower colourless acid layer separated. The remaining acid is removed by shaking with three or four successive quantities of 30 c.c. distilled water, or until the lower aqueous layer becomes coloured a strong pink. The reagent is now poured into a bottle of resistant glass together with 1/10th of its volume of a 1 per cent solution of the sodium salt and a few fragments of broken glass. The bottle should be completely filled and should be kept in a cool dark place. The reagent gradually becomes acid, but owing to the presence of the sodium salt, neutrality is restored by double decomposition. The alkali set free by the action of the moist air on the glass combines with the iodeosin, producing the corresponding salt, which, being insoluble in the aqueous ether, is left as a pink deposit on the glass. To remove the excess of iodeosin the glass is rapidly rinsed in another bath of anhydrous ether. After drying, the sides of the glass slabs are cleansed, leaving the broken surface untouched. The next process is to determine the actual amount of alkali liberated.

The glass is washed in 3 c.c. of water containing 0·1 gm. anhydrous sodium carbonate per litre, and the solution transferred to a small porcelain vessel of about 10 c.c. capacity divided into two equal chambers by a watertight porcelain partition. The dish is rinsed with a further 2 c.c. of water, and this is added to the first wash water. To the second chamber is next added 5 c.c. water, and from a pipette graduated to 1/100 c.c. a solution of sodium iodeosin containing 0·01053 gm. per litre, each cubic centimetre of which contains 0·01 mgm. free iodeosin, until the colours of the two solutions are similar.

[1] Mylius and Foerster, *Berichte der Deutsch. Chem. Ges.*, 1889, xxii. 1092; *Zeitschrift für Instrumentenkunde*, 1889, ix. 117; and *Zeitschrift für Instrumentenkunde*, 1891, xi. 311.

[2] Mylius, *Silikat Zeitsch.*, 1913, 1, 2, 25, 45, contains latest working details; Mylius, *Zeitsch. für Instrumentenkunde*, 1889, ix. 50; Mylius and Foerster, *Berichte der Deutsch. Chem. Ges.*, 1891, xxiv. 1482; Mylius, *Zeit. anorg. Chem.*, 1907, lv. 233, also 1910, lxvii. 200.

For a useful résumé of work by Mylius and Foerster see *Jena Glass*, Hovestadt, Eng. Trans. by Everett (Macmillan).

The area of the fractured surface is now calculated and the amount of iodeosin which would combine with the alkali set free from 1 sq. metre determined.

The following classification has been adopted to express the results of the tests :

Class.	Mgm. of Iodeosin per sq. metre.
H_1	0 – 5
H_2	5 – 10
H_3	10 – 20
H_4	20 – 40
H_5	40 – 80

Glasses in the first three classes are considered to withstand the action of the atmosphere and not to suffer tarnish, while those in classes 4 and 5 should not be used in exposed positions. At the present time this is the only method capable of quantitative treatment, and is therefore exclusively used for comparing and classifying optical glasses. Other methods in which the skill and judgment of the experimenter are the sole criterion are occasionally used. One of these devised by Weber consists in exposing pieces of glass to the action of hydrochloric acid fumes for twenty-four hours and then removing to the air. The less stable glasses will be found to become coated with a rime of alkaline chlorides, and from the appearance it is possible to make some form of classification. The method requires a great amount of experience and skill.

Another method due to Zschimmer, which has been recently modified by subsequent workers, seems likely to be of service in predicting the susceptibility of glasses to "filming."[1] The polished slabs of glass, together with polished slabs of transparent silica, are placed in a glass tube in a thermostat and heated to 80° C. A current of pure moist air is passed through, and after some time the heat is withdrawn, thus allowing the tube to cool; it is found on removing the large tube that some glasses are more bedewed than others, while the silica plates are usually dry. The rate of disappearance of the dew is observed by placing the samples in separate closed tubes, and on drying they are again carefully examined. In this test the good glasses are very little changed, while those which the iodeosin test has shown to be poor are frequently found to be actually pitted. Superheated steam has also been used, but the results obtained require to be viewed with caution.

(ii.) *The Testing of Chemical and other Glassware.* — This problem has been the subject of very numerous researches since the time of Scheele and Lavoisier, but owing to the fact that it is only in recent times that highly resistant glass has been manufactured, the early researches become merely of historic importance. The later workers have unfortunately not adopted a uniform method of test, and it becomes a matter of some difficulty to correlate the numerical data obtained. The conclusions, however, are on the whole fairly unanimous, and are summarised in a preceding section.

Mylius and Foerster, who were the first to study systematically the modern types of glass, were inclined to attach too great importance to the attack by water, and to consider the differences obtained with other reagents as of minor significance. Later investigations have taken more comprehensive views, and as a result a great improvement has been made in the manufacture of resistant glasses.

The action of water at ordinary temperatures, *e.g.* 20° C., also at higher temperatures, 80° C., and also boiling water and steam, both under atmospheric and high pressure, is usually determined.

The last method has recently come very much into vogue. For many glasses the test is a useful one, while for others the conditions are perhaps too severe, and the conclusions drawn may conceivably lead to erroneous results.

The action of acids, alkalies both strong and dilute, also ammonia and ammonium chloride, should also be tested, and great care should be taken that the temperatures are carefully maintained, more especially with the caustic alkalies, since the rate of attack increases very rapidly with small rises of temperature. It is impossible to give the full details of all the tests, and for fuller information the memoirs mentioned below should be consulted.

W. H. W.

Foerster, *Zeitsch. Instrumentenkunde*, 1893, xiii. 457, and *Zeitsch. analyt. Chem.*, 1894, xxxiii. 381; Cauwood English and Turner, *Journal Society of Glass Technology*, 1917, i. 153, containing résumé of previous work; Sullivan, *Journal Society Chem. Industry*, 1916, xxxv. 513. A useful bibliography of papers dealing with the stability of glass and its testing has recently been published — Turner, *Journal Society of Glass Technology*, 1917, i. 213, in which other useful papers will be found.

GLASS, OPTICAL PROPERTIES OF: Table. See "Optical Glass," § (4).

GLASS, PLATINISED

PLATINISED glass has not hitherto found much use in Applied Optics, for the following reasons. When the platinised surface is produced by deposition, or by "spluttering" methods, it is easily rubbed off, and requires protection, thus showing no advantage over silvering methods. When produced by the so-called "burning-in" process, the resulting layer is more or less granular and diffuses the light very considerably. In recent years,

[1] Elsden Roberts and Jones, *Journal Society of Glass Technology*, 1919, iii. 52.

however, Messrs. Rheinberg & Co., London, have introduced platinised and semi-platinised surface mirrors, in which the platinum is incorporated in the surface layer of the glass, so as to be irremovable, and which show no grain and are remarkably free from diffusion. They also withstand the action of all ordinary solvents, except those which attack the glass itself. Notwithstanding the fact that the intrinsic reflective power of a platinised surface is between one-third and one-fourth less than that of a perfect silvered surface, they supplant the latter, because silvered surface mirrors are liable to tarnish, are very delicate, and deteriorate fairly rapidly, whilst the platinised surface mirrors will wear, require no protection and do not suffer deterioration.

Amongst the applications of such platinised surface mirrors are reversing mirrors for process cameras, in replacement of large and expensive prisms; reflex camera mirrors, and, generally speaking, mirrors where first surface reflections only are required. They likewise find employment for dental and surgical mirrors, on account of the freedom with which they can be subjected to hot water or sterilising fluids without deterioration.

Semi-platinised mirrors find application in or with optical apparatus as light filters, since they have a neutral grey tint, and may be placed in any position where they are subjected to heat, without liability to damage or change. They are also employed in the types of optical instruments where some definite part of the light is reflected by the surface, and part transmitted through the glass.

Platinised mirrors by the Rheinberg process reflect the red end of the spectrum proportionately more than the blue end to a *very slight* extent as compared with silvered mirrors. This is easily observable by comparing the image of a person in both, the difference being just sufficient to impart to the face the appearance of being slightly ruddier and healthier. The platinum mirror, however, also reflects the ultra-violet rays just beyond the visual spectrum to a greater extent than silvered mirrors, as the latter absorb these rays more strongly. This has been demonstrated by spectrographs taken by reflection from platinised mirrors and freshly silvered surface mirrors under the same conditions.

Platinised mirrors can be used as back surface as well as front surface mirrors, but the reflection from the back surface suffers slightly in brilliancy.

The Rheinberg process employs a high temperature,[1] and its success depends on:

(1) The special coating mixtures used.

(2) Careful pyrometric regulation of the electric furnaces.

(3) The precautions used to ensure retention of the plane or curved surfaces of the glass to be platinised.

[1] See patent N. 156,472, Jan. 13, 1921.

(4) Precise regulations according to the particular composition of the glasses used.

Flint glasses are unsuitable for this process, as the platinum interacts with the lead in the glass. Whilst a large variety of glasses can be successfully platinised, the best results are obtained with those of the crown glass and plate glass description.

Fuller particulars will be found in a paper read before the Optical Society in November 1920.

J. R.

GLASS, PROPERTIES OF STRAINED. See "Glass," § (19) (ii.).

GLASS, strength of. See "Glass," § (24).

Testing of. See "Glass, Chemical Decomposition of," § (3).

GLASS ANALYSES, TABLE OF. See "Glass," § (3).

GLASS ANNEALING: the removal of strain in glass by heating and subsequent slow cooling. See also "Glass," § (19).

GLASS COLOUR, CONTROL OF, IN MANUFACTURE. See "Glass," § (16) (iv.).

GLASS DEFECTS—"CORDS." See "Striae." See also "Glass," § (16) (ii.).

GLASS WOOL, MANUFACTURE OF. See "Glass," § (18) (iii.).

GLASSES, COLOURED

THE physical nature of a glass closely resembles that of a rapidly solidified liquid, which has the power of holding metallic oxides, metals, and other matter in combination, solution, or suspension. There are many varieties of glass of very varied composition, but nearly all possess the property referred to. In Germany several scientific papers have been published dealing with the relation of glasses to effects of colour, but English contributions to the subject have been, with very few exceptions, aesthetic rather than scientific. It may, however, be claimed that in practical glass-making, especially in the recapture of the colours of mediaeval windows, English manufacturers have been quite as successful as their foreign competitors.

Variation in the chemical composition of glasses produces varied effects in their action on light. The influence of the variation on the refractive power and dispersive power of glasses is dealt with elsewhere. Variation in colour is also due to variation in chemical composition. Colour effects are caused by the power possessed by glasses of absorbing some of the constituent rays of white light and of transmitting or partly transmitting and partly scattering the remainder. Even the most colourless optical glasses show noticeable colour when a considerable thickness is traversed by

light. Glazed window glass, of good quality, appears to be colourless, but a tinge of colour shows when a sheet of similar glass is placed on white paper, and if the edge of the sheet is viewed it appears to possess a deep sea-green colour.

Coloured glasses may be transparent, translucent, or opaque: transparent, if practically all the incident rays, not absorbed, are transmitted; translucent, if the non-absorbed rays are partly transmitted and partly scattered by minute, opaque, embedded particles; opaque, if all the non-absorbed rays are reflected.

Colours are produced by introducing into glass mixtures certain ingredients, usually metallic oxides, and melting them in crucibles or tanks. The resultant colours depend on the colouring ingredient and its condition, on the ingredients of the glass mixture, on the atmosphere surrounding the molten mixture, and, in some cases, on the temperature to which it is exposed.

The *glass mixture* generally used as a basis for the production of coloured glasses is either A or B:

A—sand, carbonate of potash, and red lead.

B—sand, carbonate of soda, and carbonate of lime.

§ (1) MANUFACTURING PRACTICE.—The following is a record of some results obtained on a fairly large scale in a manufactory:

Copper, used in the form of *copper scales*:[1] with A mixture, 0·8 per cent, light blue colour, like that of cupric sulphate; with A mixture, 5 per cent, dark green colour; with B mixture, 0·2 per cent, bright blue; with a sand-potash-lime mixture, to which stannic oxide, red tartar,[2] 0·3 per cent copper scales, and 0·1 per cent of ferric oxide have been added—a colourless glass which gradually becomes very dark ruby when reheated. If the atmosphere surrounding the molten glass has not had a sufficiently reducing effect, it appears, when cold, to be rusty, and transmits a blue colour.

A similar mixture, but with carbonate of soda substituted for carbonate of potash, in a strongly reducing atmosphere, develops visible crystals of metallic copper, and a glass closely resembling the mineral aventurine. The crucible containing the molten mixture must be cooled extremely slowly.

Gold, dissolved in aqua regia: with a sand-potash-lead mixture to which crushed metallic antimony, stannic oxide, antimony trioxide, charcoal, and 0·03 per cent gold have been added — a colourless glass which gradually develops a dark ruby colour when reheated. If the atmosphere during melting has not had a sufficiently reducing effect, the glass appears to be rusty-red and transmits a violet-blue colour.

Iron, usually used in the form of ferric oxide: ferric oxide with A mixture, 5 per cent, a rich yellow colour; 1 per cent of black oxide of manganese is usually added to give the yellow an amber tint. Iron scales[3] with B mixture, 0·1 per cent, a pale blue colour. With a larger proportion of iron scales, a grey-green.

Ordinary bottle-green colour is usually obtained by adding a small proportion of cobalt oxide to from 3 to 5 per cent of ferric oxide.

Manganese, used as manganese dioxide: with either A or B mixture, about 1 per cent, a violet colour. The oxide is used as a decolorant for A glasses. It acts partly by converting any iron in the glass mixture into ferric oxide, and partly by producing a tint of colour complementary to the green due to ferrous oxide.

The oxide used as a decolorant with B glasses, *e.g.* ordinary window glass, although the glass may be colourless when first made, develops a pink colour in the glass after long exposure to light. The addition of a colouring ingredient to neutralise colour in a glass mixture reduces the transparency of the glass.

Cobalt, used as black oxide: with A or B mixture, 0·4 to 1 per cent, a bright purple-blue, less purple with B than A. Cobalt glasses with large proportion of the oxide, or in considerable thickness, transmit a ruby colour. The effect can be neutralised by the addition of chromic oxide or cupric oxide.

Nickel, used as nickel monoxide: with A mixture, 0·2 per cent, a strong violet colour; with B mixture, a brown colour.

Sir Herbert Jackson, F.R.S., has found that the same proportion of the oxide with a lithia glass gives a yellowish-brown colour.

Chromium, used as sesquioxide: 0·4 per cent, with both A and B mixtures, gives a greenish-yellow colour.

Uranium, used as uranate of soda: 0·5 per cent with A or B mixture, fluorescent yellow colour. Colour and fluorescence better developed with B than with A mixture.

Carbon, used as finely divided wood charcoal: with B mixture, 0·02 per cent, a clear yellow colour. As the proportion of carbon is increased the colour passes from yellow to brown, brown to black.

Arsenic trioxide: with A mixture, 5 per cent, gives an opaque white enamel; a thin film of the enamel transmits a strong amber-red colour.

Fluor spar: with A mixture, 4 per cent, gives a deep opal effect.

[1] Copper scales = crude red cuprous oxide, Cu_2O.
[2] Red tartar = crude bitartrate of potash.
[3] Iron scales: a thin black film formed when iron is heated in air = Fe_3O_4 ferroso-ferric oxide.

Silver: used in the form of oxide, as a pigment. It is mixed with some finely divided difficultly fusible material, such as kaolin, applied to the surface of a colourless glass, and strongly heated. It gives a yellow transparent colour, with a slight bluish lustre. The colour is best developed on a soda-lime glass, of which phosphate of lime is an ingredient.

§ (2) Experimental Meltings.—*Platinum black*: with A mixture, 0·3 per cent, is diffused through the glass, giving a grey translucent effect.

Oxide of iridium: with A mixture, 3 per cent, gives a dense black.

Carbonate of thallium: with A mixture, 6 per cent, gives a greenish-yellow colour.

Antimony trioxide: with A mixture, 3 per cent, gives a pale translucent yellow.

Vanadium, as pentoxide: with A mixture, a yellow-green; with B mixture, a blue-green.

Selenium, as a selenate: with A mixture, a pink amber.

§ (3) Crookes' Glasses.—Sir William Crookes' prolonged researches[1] to discover satisfactory eye-preserving glasses for spectacles throw light not only on the effects produced by special ingredients introduced into glasses in cutting off ultra-violet and heat rays, but also in the production of effects of colour. The base used for all of his experiments was a soda-lime glass, and the thickness of the plates of glass, when prepared for use, was 2 mm. The results of his experiments with regard to colour are as follows:

Cerium nitrate 17 per cent, pale reddish-amber tint; chromic oxide 1 per cent, green; copper sulphate about 2 per cent, blue; ferrous oxalate 10 per cent, a smoky green colour; ferric oxide 2 per cent, yellow; ferroso-ferric oxide, 2·85 per cent, with the addition of carbon 0·35 per cent, pale blue. Manganese gives a reddish-purple colour, cobalt sulphate a rich blue, nickel sulphate a brown, cobalt mixed with nickel a neutral grey, praseodymium a greenish-yellow, neodymium a lilac.

The conclusions arrived at with regard to eye preservation were that a glass containing cerium is most effective in cutting off injurious ultra-violet rays, and a glass containing iron, in a ferrous or metallic condition, is most effective in cutting off heat rays. A glass composed of 83 per cent of fused soda glass and of 17 per cent cerium nitrate was found to be practically opaque to ultra-violet radiation, the limit being $\lambda 3650$; a glass composed of 90 per cent raw soda glass mixture, and 10 per cent ferrous oxalate, with small additions of red tartar and wood charcoal, gave a sage-green glass which cut off 98 per cent of heat radiation; whilst a glass composed of 96·80 fused soda glass, 2·85 ferroso-ferric oxide, and 0·35 carbon had a pale blue colour, and cut off 96 per cent of heat radiation.

[1] *Phil. Trans. R. Society*, Series A, 1913, ccxiv.

§ (4) Absorption of Light by Coloured Glasses.—R. Zsigmondy's papers on the absorption of light by coloured glasses,[2] and on the use of coloured glasses for scientific and technical purposes,[3] have some bearing on the subject of this article. He rightly insists that the colour of a glass depends not only on the colouring material, but on the glass mixture or batch to which it is added and on the atmosphere in which it is melted. As examples he gives (1) the brown- and violet-coloured glasses produced, respectively, by a soda-lime and potash-lead glass mixture with the same proportion of the same oxide of nickel; and (2) the change in colour from yellow to yellow-green, and from yellow-green to blue-green, of a glass mixture containing ferric oxide when, in the process of melting, an oxidising atmosphere changes to a reducing one. He calls attention to the action of manganese in reducing the green part of the spectrum and to the increase of the red part of the spectrum when a small proportion of manganic oxide is added to the ferric oxide of a mixture for yellow glass. In dealing with copper he found that if the proportion of cupric oxide be increased from 2 per cent, which gives a blue glass, to three or five times that value, green glasses are obtained differing widely in their absorption from that of a copper-blue glass. In his first research the following bases and glass mixtures were used; borax, sodium silicate, potassium silicate, lead silicate, soda-lime glass, potash-lime glass, soda-lead-lime glass, potash-lead glass, soda-zinc-lime glass, soda-borosilicate glass, and barium-borosilicate glass. The colouring agents were calculated as oxides, and the proportions ranged, except for the green-copper glasses, from 0·1 per cent to 2 per cent. The mixtures were melted and stirred in an oxidising atmosphere; the molten glasses were poured, and the plates, thus formed, after annealing, were cut and polished for examination. A "Glan" spectrophotometer was used, and the coefficients of extinction, having been determined, were used for the construction of representative curves, of which illustrations are given. It is regrettable that in this research the ordinary soda-lime and potash-lead glasses were so rarely used that it is difficult to compare their effects. The objects of the second research on the use of coloured glasses for scientific and technical purposes were (1) to provide a ray-filter, which would transmit one part of the spectrum whilst absorbing the remainder; (2) to obtain light filters for three-colour photography;

[2] *Ann. d. Phys.*, 1901, iv. 60.
[3] *Zeitschr. f. Instrumen.*, 1901, xxi. 97.

and (3) to find a glass to absorb all parts of the spectrum equally. A Pulfrich comparison spectroscope was used for examining the glasses, and the following is an abridged table of the results. It is unfortunate that in the account of this research the compositions of the glass mixtures are not given.

Glass.	Colour.	Spectral Rays transmitted.
Copper-ruby glass	Deep red	Only red.
Gold-ruby glass	Red	Red, yellow; in thin sample, blue and violet.
Uranium glass	Bright yellow	Red, yellow, green; in thin sample, blue.
Nickel glass, soda base	Yellowish-brown	Red-yellow, traces of green and blue.
Nickel glass, potash base	Dark violet	Violet (G – H), extreme red.
Copper glass, blue	Blue, like copper sulphate	Green, blue, violet.
Copper glass, green	Green	Green, yellow, traces of red and blue.
Chrome glass	Yellowish-green	Yellowish-green, traces of red.
Cobalt glass	Blue	Blue, violet, extreme red.
Manganese glass	Dark violet	Violet (G – H), extreme red.
Smoke-grey glass	Grey	Whole spectrum weakened.

The following results with specimens of mediaeval glass were obtained by the author of this paper:

Glass.	Colour	Spectral Rays transmitted.
Copper-ruby of thirteenth century	Deep red	Red, bluish-green and trace of yellow.
Copper-ruby of fourteenth century	Deep red	Red and some blue.
Blue of thirteenth century	Greyish-blue	Blue, violet, traces of red and yellow, but no green.
Carbon-yellow of thirteenth century	Pale yellow	Passes all rays except blue, violet, and extreme red.

§ (5) Suggested Explanations of Effects of Colour of Transparent Glasses.—The behaviour of metallic oxides introduced into a glass seems to correspond exactly with their behaviour when added to an acid solution. Thus cupric oxide, dissolved in hydrochloric acid of moderate concentration, gives a green solution of cupric chloride, which contains in addition to the undissociated salt a number of blue cupric and colourless chlorine ions. If this solution is largely diluted with water the ionisation is increased, and the colour of the cupric ions predominates over that of the undissociated residue, the solution becoming blue. Precisely the same result is obtained when cupric oxide is added to glass, which it must be remembered always contains a large excess of the acidic oxide, silica. When present in comparatively large quantities the cupric oxide gives a green glass, but if only a minute amount is added a blue glass results. Similar resemblances can be traced in the behaviour of other metallic oxides. It is probable then that ionic dissociation of the metallic silicates takes place on melting. It might be thought that the existence of this ionisation would render a coloured glass a conductor of electricity. This, however, cannot be the case since the ions in a solid have no mobility, with the result that a blue cupric glass is as good an insulator as a dry crystal of copper sulphate, in which ions are present in large numbers, but without the power of moving through the crystal.

There is another group of coloured glasses in which the origin of the colour is entirely different, examples of which are the ruby glasses made by the formation of reduced gold and cuprous oxide in the mass of the glass. Here again we may see a close resemblance between the behaviour of a salt solution and a glass. Carey Lea in 1884 showed that if a very dilute solution of gold were treated with a reducing agent a bright red solution, apparently quite transparent, was produced. With increased concentration a purple solution was formed, and in still stronger solutions a yellow colour resulted. These solutions, so-called colloidal, are now known to be not solutions at all, but suspensions of particles so small that their presence can only be demonstrated by the ultra-microscope.[1] In gold and cuprous oxide ruby glasses this method of viewing them renders it quite certain that they owe their colour to the presence of fine particles, the size of which, by their action on the light of different wave-lengths, determines the colour of the transmitted and reflected light. It may also be assumed that the yellow and brown colour of carbon glass, for which many different causes have been suggested, is really due to ultra-

[1] See also paper by J. C. Maxwell Garnett, *Phil. Trans. Royal Society*, A, 1904.

microscopic particles of silicon, comparable to the particles of metallic gold and cuprous oxide, which are the cause of colour, respectively, in gold-ruby and copper-ruby glass.

H. J. P.

GONIOMETRY

§ (1) Spectrometer Methods.—The measurement of the angles between the flat surfaces of prisms is one which for many purposes requires to be performed with the greatest possible accuracy. The refracting angle of a prism to be used for refractive index measurements, for example, must be known to within about one second if the index measurements are to be correct to the fifth decimal place; and some of the angles of prisms used in rangefinders, dial sights, and similar instruments must approximate to specified values to within about the same amount. We shall first treat the general methods, and then refer to special methods suitable for special cases. For the general methods a spectrometer is required. The features of such an instrument adapted to the most precise work are discussed in another article [1] and the various adjustments described.

The two best-known methods of measuring an angle on the spectrometer are the following:

Method 1.—The prism is placed on the spectrometer table with the vertex of the angle in question over the centre (*Fig.* 1 (*a*)), and the table is adjusted until the edge of the prism is parallel to the axis of rotation.[1] The parallel beam from the collimator is divided into two by reflection at the faces of the prism. If the telescope is placed in the position T_1, one of the reflected images can be brought to the cross-lines, and in the position T_2 the other. The angle between the two positions of the telescope is twice the angle of the prism.

Method 2.—The prism is placed with its centre over the centre of the table and rotated so that the reflected image from one face is on the cross-lines of the telescope T (*Fig.* 1 (*b*)). The table is then turned until the reflection from the other face of the required angle is on the cross-lines. The second face is now in exactly the position previously occupied by the first; and the angle through which the table has been turned is 180° minus the angle of the prism.

[1] "Spectroscopes and Refractometers," § (8).

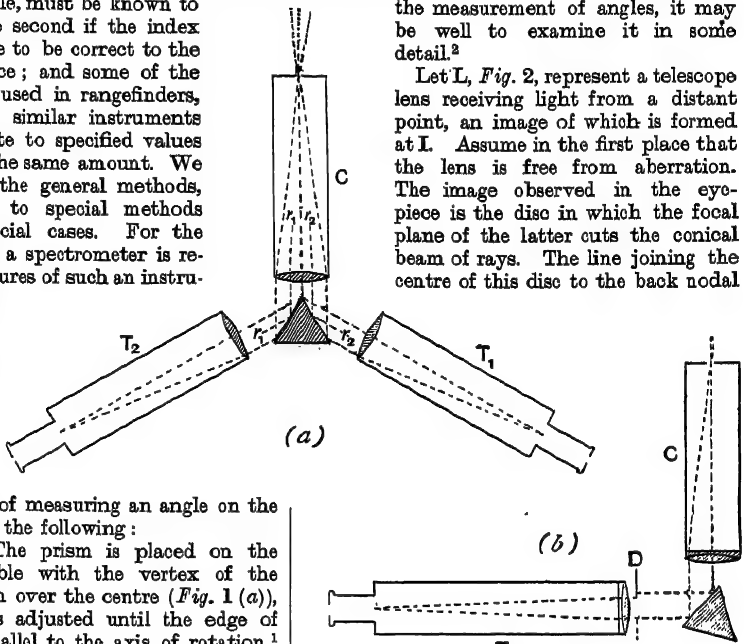


Fig. 1.

For work of only moderate accuracy—say 10 seconds or so—there is little to choose between these methods; but when high accuracy is aimed at Method 1 is quite useless. This arises from the fact that in practice it is impossible, except by accident, to obtain perfect collimation or perfect focussing. There is always a range, known as the depth of focus, within which the definition of an optical instrument appears uniformly good, and there is consequently a certain degree of latitude in all adjustments of collimation or focus. As this error vitiates quite a large proportion of the methods which have been published for the measurement of angles, it may be well to examine it in some detail.[2]

Let L, *Fig.* 2, represent a telescope lens receiving light from a distant point, an image of which is formed at I. Assume in the first place that the lens is free from aberration. The image observed in the eyepiece is the disc in which the focal plane of the latter cuts the conical beam of rays. The line joining the centre of this disc to the back nodal point of the objective is the direction of the incident beam as indicated by the telescope. Clearly, provided the aperture is symmetrical about the back nodal point of the lens, the apparent direction is not affected by errors of focus; for if the plane of the cross-wires is not accurately at I but is at I′, the centre of the disc of confusion is still on the line IN. If, however, one half of the object-glass is covered by a screen, as in *Fig.* 2 (*b*), the centroid of the disc of confusion at I′ is below the line IN, while at I″ it is above it. In the first case the cross-lines will be set to the left of their proper position, while in the second case

[2] See also Guild, *Proc. Phys. Soc.*, 1916, xxviii. 242.

they will be set too far to the right. If the other half of the objective be employed the errors will be reversed. In general, unless the centroid of the section of the beam by the principal plane of the lens coincides with the nodal point, the centroid of the image will be displaced from its true position unless the focus is absolutely exact. If then the focus is not quite accurate, and different parts of the aperture are employed for different component measurements of a determination, the results will be in error. In the presence of spherical aberration there is no position of focus at all for which the centroid of the image is unaffected by cutting off part of the aperture unsymmetrically. In *Fig.* 2 (c) the case of a lens with under-corrected spherical aberration is shown (much exaggerated). f_a is the focus for rays near the axis, and f_p for those near the periphery. F is the position of best focus when the whole aperture is used, while if either half is used separately, the disc of least confusion would be displaced appreciably towards F′ or F″. The effect on the displacement of the image from its true position on the line NA for different positions of the focal plane of the eyepiece is easily followed from the diagram.

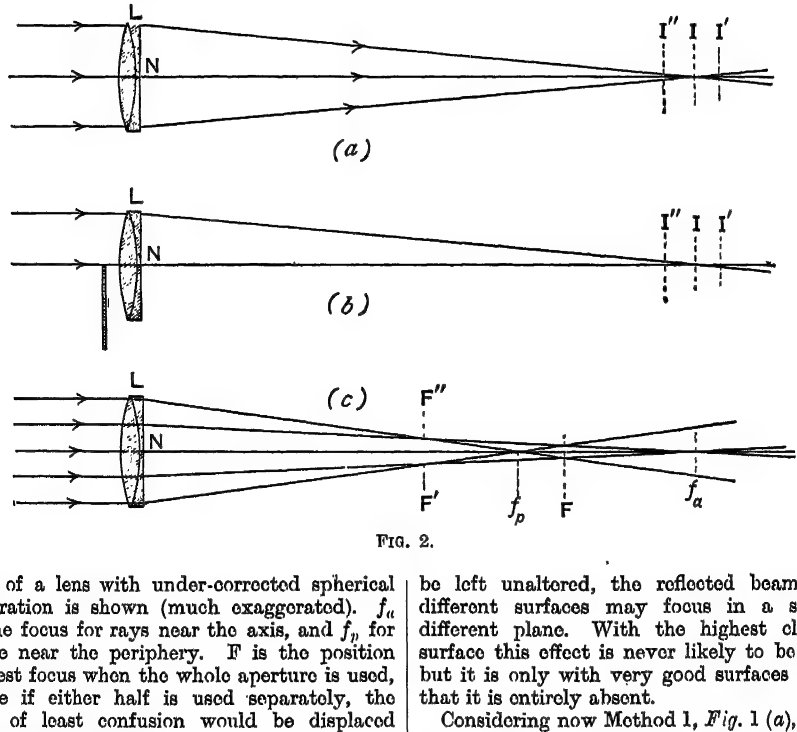


FIG. 2.

In practice, if an object be sighted with a telescope with partially obscured aperture, the movement of the image across the field as the focus is adjusted is very easily observed. Within a certain range there is no means of judging when to stop. Even with very sharp focussing it is easy for errors of several seconds to be made, while under conditions by no means abnormal in laboratory work errors as great as half a minute may arise. It is obvious, of course, that the error is in a plane perpendicular to the line of division of the objective, and that measurements in the horizontal plane will not be affected, provided the aperture is symmetrical with respect to the vertical diameter of the lens.

Thus an essential condition to the accuracy of measurements of angles with a telescope is that the aperture employed should be symmetrical with respect to that diameter of the lens which is perpendicular to the plane of measurement. It is not safe to use an excentric aperture, even though the same region be used for all the measurements; for, although the focal adjustment of the telescope be left unaltered, the reflected beam from different surfaces may focus in a slightly different plane. With the highest class of surface this effect is never likely to be great, but it is only with very good surfaces indeed that it is entirely absent.

Considering now Method 1, *Fig.* 1 (*a*), in the light of these considerations, it is evident that in changing from position T_1 to T_2 the full effect of errors of focus or spherical aberration will be felt, since opposite halves of the objective are used for the two measurements. If to obviate this we place the prism in such a position that the aperture of the telescope is symmetrically filled, the ray from the vertex no longer passes through the nodal point, and the directions actually measured by the telescope are those of the rays r_1r_1 and r_2r_2. There is therefore liability to error on account of imperfect focus of the collimator; since, unless collimation is perfect, r_1 and r_2

do not emerge quite parallel, and the angle between the reflected rays is not equal to twice the prism angle. It is therefore immaterial where the prism is placed, when inaccuracies due not only to the collimation but to the focussing of the telescope as well are taken into consideration. It is impossible to arrange the method to be simultaneously free from both effects, and it cannot be used for accurate work.

Method 2, on the other hand, is free from this defect. By means of an iris diaphragm D,[1] which is closed down until a circular aperture completely filled with light is obtained, exact symmetry may be assured. As it is desirable to use the biggest available aperture, the telescope axis should be rotated, if possible, out of its usual position (directed to the centre) and directed to the middle of the side of the prism. The fact that the collimator is not symmetrically employed is clearly of no consequence under the conditions of the experiment, provided the prism is mounted with the bisector of the angle over the axis of rotation, so that the two surfaces are in the same plane when the light is reflected from them.

This is the most convenient method for general use, provided the rotation of the table can be determined with the same accuracy as that of the telescope. If, however, as is too frequently the case, no provision exists for the *accurate* measurement of table rotations, it is necessary to resort to the next method, which involves the use of an auto-collimating eyepiece.

Method 3.—The telescope is fitted with an eyepiece arranged so that the cross-lines may be illuminated from behind and the telescope itself may be used as a collimator. When its axis is approximately normal to a plane surface, a reflected image of the cross-lines will be seen in the focal plane. When these are accurately set to coincide with the actual cross-lines, the axis of the telescope is normal to the surface. The determination of an angle consists, therefore, in setting the telescope normal first to one surface and then to the other. The angle between these positions is equal to 180° minus the angle of the prism.

There are several forms of auto-collimating eyepiece, but not all of them are useful. Those forms in which the reflector is behind any of the lenses of the eyepiece are most troublesome to use, on account of the flooding of the eye with light reflected from the surfaces of these lenses. The most satisfactory form of auto-collimating eyepiece is that due to Abbe. It is illustrated in *Fig.* 3. A small slab of glass, *p*, bevelled at one end to form a 45° reflecting prism, is mounted behind part of the cross-lines. This prism, when illuminated by a lamp placed above it, acts as a bright background to the cross-lines in front of it. When the telescope is approximately normal to a reflecting surface the appearance of the field may resemble *Fig.* 3 (*c*). The whole field will probably be faintly illuminated by the

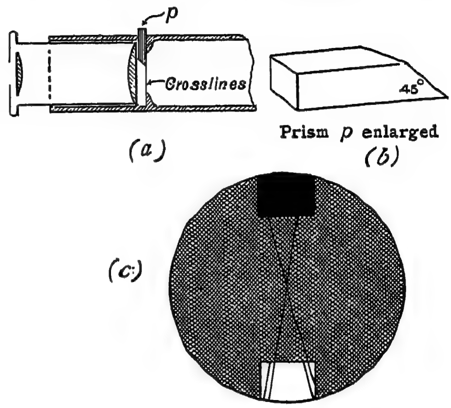


FIG. 3.

general light of the room, except for the area occupied by the back of the prism *p*, which will be quite dark. At the opposite side of the field to this will be the brightly illuminated image of the face of *p*, with the image of the cross-lines showing up against it. When the normality is exact, both horizontally and vertically, the actual cross-lines and those of the image will coincide. It improves the accuracy of setting considerably if the axis of the telescope is *slightly* inclined in the vertical direction. In that case the image lines are a little wider or closer than the actual cross-lines at the same distance from the centre of the field, and instead of a somewhat insensitive coincidence setting a symmetrical setting is made with the image lines a little outside or inside the others.

This is a very fine eyepiece to work with. Its only drawback is the difficulty of using a high-power ocular, on account of the distance between the cross-lines and the field lens necessitated by the insertion of the illuminating prism.

Another arrangement which may be employed is not a true auto-collimating eyepiece but serves the same purpose. It is shown in *Fig.* 4. A side tube is attached to the telescope just in front of the eyepiece, and serves to carry a short draw-tube at the end of which can be attached a pinhole, a fine slit, or a graduated scale, or whatever type of object may be most convenient for the purpose on hand. A semi-transparent reflector, which may be a microscope cover-slip thinly coated

[1] It is a great convenience to have iris diaphragms fitted to spectrometer telescopes and collimators.

with silver or platinum,[1] is mounted at R by some suitable means. When the telescope is nearly normal to a reflecting surface, an image of P is seen in the field of the eyepiece. For use in Method 3, above, the most suitable object is a moderate-sized hole with a vertical fibre and two horizontal fibres. These latter should be about $\frac{1}{2}$ mm. apart, and settings are made on the vertical fibre half-way between them.

In carrying out measurements by Methods

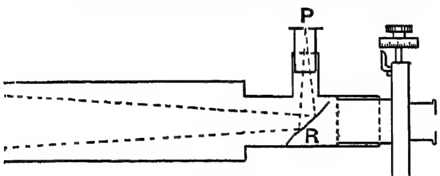


FIG. 4.

2 or 3, every care must be taken to eliminate all the sources of error to which even the best spectrometers are liable. The article on "Spectroscopes and Refractometers" should be consulted for a discussion of these errors. In Method 2 it will be found advantageous to employ a broad slit and fibre, as described in the article just referred to, rather than a narrow slit. It is advantageous to employ a light green filter in front of the light source in all goniometric measurements as this eliminates chromatic aberration and improves definition.

§ (2) Method of Substitution.—With an exceptionally fine spectrometer it is possible to measure angles by Methods 2 or 3 to within a second, provided extreme care is taken. Spectrometers of this quality are few; and it is not every observer who is fortunate enough to have access to one.

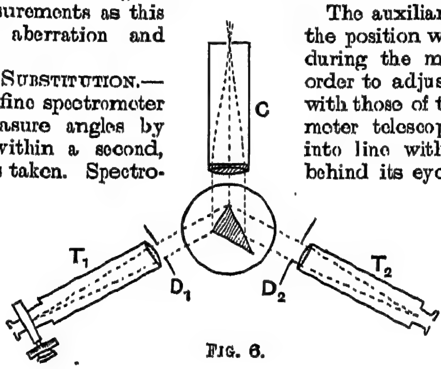


FIG. 6.

The overwhelming majority of angles in optical prisms are sub-multiples of 180°; and of these the great bulk are 45°, 60°, or 90°. For such angles special methods of high accuracy can be employed which do not involve the measurement of an angle on the spectrometer at all. It is therefore possible to have standard angles of these values, and the actual testing of an unknown angle is reduced to a measurement of the small difference between it and the corresponding standard. We shall first describe the method of substitution, and then refer to the methods of obtaining the standards. If a spectrometer is available, the only additional apparatus required is a second telescope on a separate stand. This telescope should be similar to that of the spectrometer, but only one of them requires to have a micrometer eyepiece.

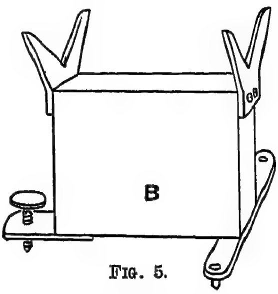


FIG. 5.

Such an auxiliary telescope is of great use in many ways,[2] and there should be one associated with every spectrometer. A cheaply constructed stand of perfect rigidity can easily be made as in *Fig.* 5. The body, B, is a slab of hard wood which carries metal Vs in which the telescope rests. The method is carried out as follows:

The auxiliary telescope T_2 is placed in the position which it is going to occupy during the measurements, *Fig.* 6. In order to adjust its axis to be coplanar with those of the collimator and spectrometer telescope, the latter is rotated into line with T_2 and a lamp placed behind its eyepiece. The cross-lines of T_1 are then visible in the field of T_2, and the latter is adjusted until both sets of cross-wires are at the same height. This procedure assumes of course that the axes of T_1 and C are already in adjustment with respect to the centre. If this is not so the complete adjustment must be made for all three at once, as described in "Spectroscopes and Refractometers," § (8).

One of the nearly equal angles which have to be compared is placed on the table so that light from one of its faces is directed into the auxiliary telescope T_2. The other telescope T_1 is brought to receive the image from the other face. The table is levelled until both images are at the correct height with respect to the cross-lines. This is most rapidly done if one of the faces is perpendicular to the line joining

[1] A small piece of glass may be very conveniently semi-platinised by using the preparation known as "liquid platinum." This is painted on the surface *very thinly* and the liquid evaporated off over a hot-plate. To get a semi-transparent film, the painted surface should appear of a light amber colour when held between the eye and a light.

[2] See "Spectroscopes and Refractometers," §§ (8), (10).

two of the adjustment screws, this face being adjusted last.

When the preliminary adjustments have been made, the image in T_2 is brought to the cross-lines by means of the tangent screw which rotates the table. The micrometer eyepiece of T_1 is employed to set on the other image. This angle is then removed, and the other is laid as nearly in the same position on the table as possible and levelled in the same way as the other. The image in T_2 is again set on the cross-lines by rotation of the table. It is evident that if the angle were precisely equal to the previous one, the image in T_1 would now also be on the cross-lines. If the angles are not quite equal, the image in T_1 will be displaced by twice their difference, which is measurable with considerable precision on the micrometer. The settings *both* on T_2 and T_1 should be repeated several times.

The determination takes very little time, and with telescopes of about 30 cm. focal length should be accurate to less than 1 second. The method is clearly applicable to any size of angle for which there is a standard available.

The precautions concerning the symmetrical filling of the telescope apertures and the use of the diaphragms D_1 and D_2 must be observed, lest the focus of the reflected beams should differ slightly with the two prisms. It is also essential that the spectrometer and auxiliary telescope should stand on a more rigid support than the ordinary laboratory table, otherwise relative movements of the two will take place as the observer moves about. A convenient plan is to set them on a slab of slate or stout plate glass, which is raised off the surface of the table by three pieces of indiarubber or thick card. The auxiliary telescope then retains its position as accurately as if it were part of the spectrometer.

If much measurement of angles has to be done, it is advantageous to dispense with the spectrometer altogether and build up a simple apparatus specially for the purpose. Both telescopes and the collimator can be supported by separate stands similar to *Fig.* 5. The only additional requisite is a prism table, capable of a small amount of rotation by means of a tangent screw, and having the usual adjustments for levelling the prism. All the apparatus should stand on a slab of slate or thick glass as described above. To adjust the axes of the telescopes to be coplanar, the following procedure should be followed. Place first one and then the other telescope in line with the collimator, and adjust so that the centre of the slit is at the height of the cross-wires. If the axis of the collimator were parallel to the supporting slab to begin with, the telescope axes would now also be parallel to it, and the three axes would be coplanar in any position. If however the collimator was tilted, say downwards, each telescope will be tilted upwards. Place the telescopes in line with each other, and illuminate the cross-lines of one by means of a lamp behind the eyepiece. On looking into the other, both sets of cross-lines will be seen, differing in height by twice the amount that the axes are inclined to the base. Bring them together, making half the adjustment with each telescope. They should now both be parallel to the base. Place one of them in line with the collimator again and adjust the latter. If the initial error was considerable the adjustment should be repeated. When all three are so adjusted that their axes are parallel when taken two by two in this way, they are also parallel to the base-plate and so are coplanar in any position. They can then be set up in the relative positions of *Fig.* 6, and the prism set on the table and adjusted as already described. If it is not convenient to provide the table with the necessary fine motion for bringing the image to the cross-lines of T_2, this can be dispensed with by providing a micrometer eyepiece for T_2 as well as for T_1. A comparison is then made by placing the two prisms as nearly as possible in the same position on the table and making settings with both micrometers.

With this simple apparatus, small differences of angle can be measured with exceedingly high accuracy.

A variant of the method, which eliminates the use of a collimator, employs auto-collimating telescopes. The type of *Fig.* 4 should be used, as this is convenient for use with micrometer eyepieces. The telescopes in this case are set normal to the faces, the difference in the micrometer readings when one angle is replaced by the other being equal to twice the difference in the angles as before.

§ (3) Determination of Standards.—The substitution method of § (2) gives the easiest and most accurate method of determining an angle, provided a standard angle is available from which it only differs by a small amount. The method itself enables us to determine standards for the most usual cases which occur in practical optics. *For a* 60° *standard*, for instance, a prism with all three surfaces polished is required. By taking the differences of the three angles in pairs and assuming the sum to be 180°, each angle is obtained absolutely.

For a 90° *standard* a square prism should be employed. Here again the substitution method provides the differences of the angles in pairs, and the sum, 360°, being known, the value of each angle is obtained.

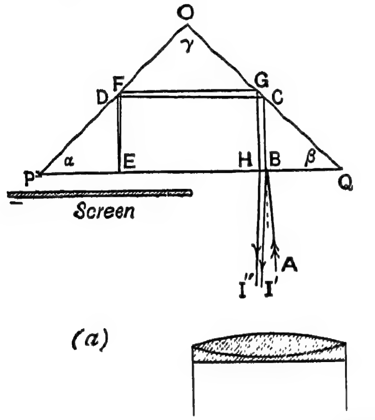
For a 45° *standard*, which may most conveniently take the form of an ordinary total reflecting prism, the difference in the 45° angle is measured by the substitution method,

and the right angle compared with the 90° standard. Thus the actual values of the 45° angles can be obtained.

For a 30° *standard* a prism of 30°, 60°, and 90° should be used. The last two angles are directly compared with the appropriate standards, 180° minus their sum giving the value of the remaining angle.

Thus the substitution method is self-contained, inasmuch as it can be employed to determine absolutely the standards required for ordinary test work.

§ (4) Auto-collimation Methods.—There are many methods of determining the commonly occurring angles in which the auto-collimating telescope is employed.[1] The most suitable telescope to employ is a small astronomical telescope of about $2\frac{1}{2}$ or 3 inch aperture, fitted with a micrometer eyepiece and the illuminating device of *Fig.* 4. The general method will be understood from *Fig.* 7 (a), which shows the arrangement for measuring the right angle of a 45° prism. The prism is placed in the position shown relative to the auto-collimating telescope, from which an incident beam AB, approximately normal to the hypotenuse of the prism, emerges. Two images will be seen in the field of the eyepiece, one due to rays such as BI′ reflected from the front face, the other due to rays which have passed through the prism by the path BCDE, being reflected approximately normally and returned along the path EFGHI″. The angular displacement of these two images depends only on the value of the right angle. The formula is easily deduced directly; but the following general method of treatment will be found useful. T. Smith[2] pointed out the great advantage, when dealing with the passage of rays through prisms, of "rectifying" the path of the ray. If one looks into a prism in any direction, he appears to be looking straight through a series of chambers which are reflections of the prism itself at the various surfaces encountered by the light rays. If these reflected prisms are drawn until a surface parallel to the front surface is reached, the path of any ray between these two surfaces is a straight line. When this is done for the cases which we are about to consider, it is found that they all reduce to the case of reflection from the front and back surface of a slightly prismatic plate. The deviation between such rays is $2\mu\theta$, μ being the refractive index and θ the inclination of the surfaces of the equivalent plate. The internally reflected ray is deviated away from the vertex of θ. The prism is developed in this way in *Fig.* 7 (b), the path of the ray within the prism being BEH. Let $\gamma = 90° + \delta$, δ being the error of the right angle. Then $\alpha + \beta = 90° - \delta$. The inclination between PQ and P′Q′ $= 180° - (P\hat{Q}P' + QP'Q') = 180° - 2(\alpha + \beta) = 2\delta$, the convergence being towards the left. The inclination of HI″ to BI_1 is therefore $4\mu\delta$. If the inclination were in the other direction, it would indicate that γ was less than 90°.



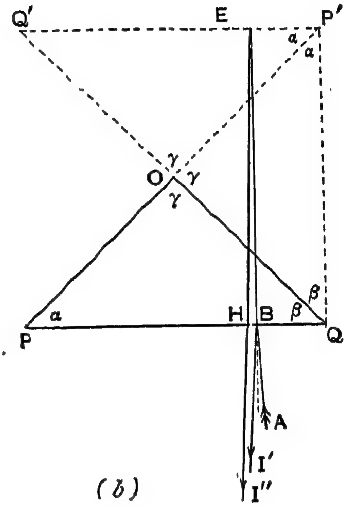


Fig. 7.

In *Fig.* 8 a number of other cases are drawn. A few words about each will suffice.

(a) To measure the difference between the 45° angles of a right-angled prism. $\theta = \alpha - \beta$. Deviation $= 2\mu(\alpha - \beta)$. In direction shown if $\alpha > \beta$.

[1] See Guild, *Proc. Phys. Soc.*, 1916, xxviii. 242; S. D. Chalmers and H. S. Ryland, *Trans. Opt. Soc.*, 1904–1905, p. 34.

[2] *Trans. Opt. Soc.*, 1918, xix. 120.

(*b*) Difference of two angles of equilateral prism. Deviation $=2\mu(\alpha-\beta)$.

(*c*) Prism of 90°, 60°, and 30°. Light incident on shortest face. $\theta=\pi-(3\beta+\gamma), =3p+q$ if $\gamma=90°-q$, and $\beta=30°-p$. Deviation $=2\mu(3p+q)$.

In general for a prism of this character in which one angle is $90°-q$ and another is $90°/n-p$, p and q being small errors, the divergence is $2\mu(np+q)$.

The defect of the auto-collimating methods for precise work is that the beam which traverses the interior of the prism is almost always thrown out of focus on account of slight want of homogeneity of the glass, and the accuracy of which the method would otherwise be capable is thus discounted. A second serious defect is that if the prism has pyramidal error (see next paragraph) the two images are at different heights in the field. This complicates the measurement very considerably. The difficulty is sometimes got over by modifications of the arrangements described, in which the condition of symmetrical use of the telescope objective is violated. In general practice therefore, in accuracy as well as convenience, the auto-collimating methods are inferior to the method of substitution.

§ (5) Defects of Prisms.—In the treatment of the previous paragraphs we have assumed perfect prisms, *i.e.* perfectly plane surfaces with the three edges parallel. In practice the surfaces are rarely absolutely flat and the edges generally converge slightly, *i.e.* the prism is part of an elongated pyramid. Clearly the measurements will only be of the highest accuracy provided the surfaces are flat enough for such accuracy to have any meaning. If the different parts of the surfaces vary in direction by several seconds, it is meaningless to attribute a more closely specified value to the angle between them. Prism surfaces are frequently convex or concave, being really long radius curves. This in practice may give rise to unexpected discrepancies between results obtained by different methods, in which, owing to the exigencies of the apparatus, different portions of the surface may be employed.

(i.) *Pyramidal Error.*—When the three edges are not parallel the prism is said to have "pyramidal error." With the exception of the general methods of § (1), the methods of precise goniometry which we have considered involve the assumption of the value of the sum of the angles of a prism. In the presence

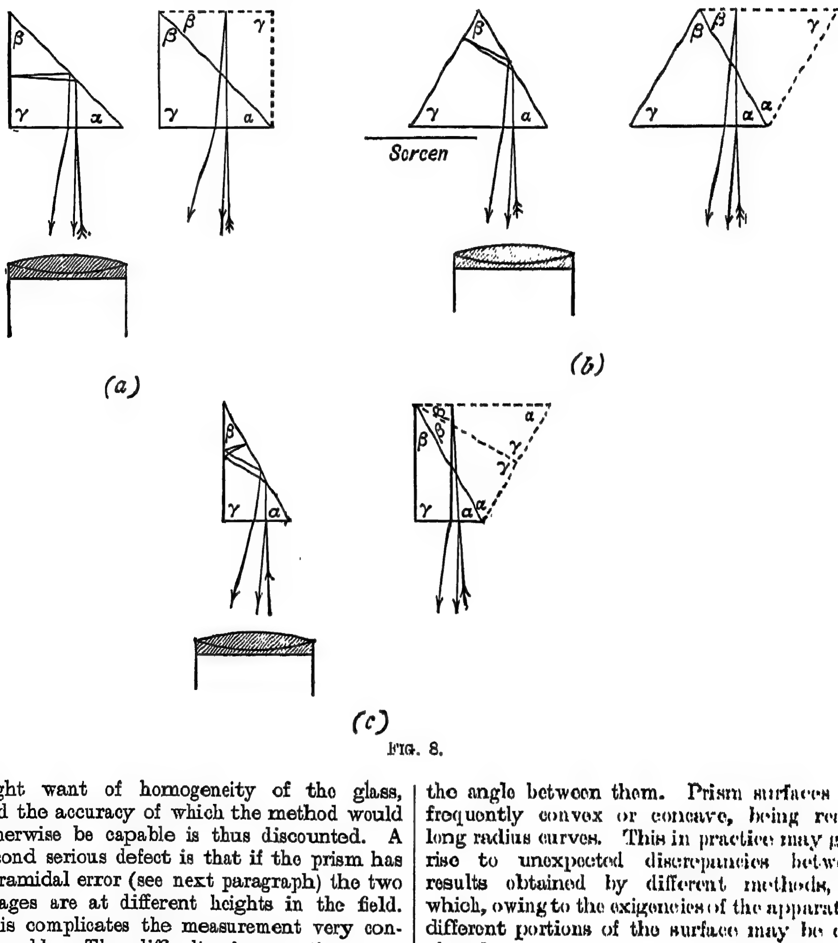


Fig. 8.

of pyramidal error this is not exactly 180°, since the angles measured are the actual angles between the faces, and these are only equal to the angles of a plane triangle when the normals are coplanar. The methods of defining pyramidal error vary considerably, but we shall define it here as the angle between one of the surfaces and the opposite edge. In the case of an equilateral prism the value is the same whichever face is considered, but in other cases it is not so. For any isosceles prism it is convenient to take as the pyramidal error the angle between the unique face and the opposite edge.

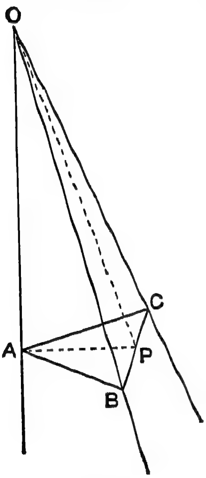


FIG. 9.

Let OBC, *Fig.* 9, be the plane of the face in question, and OA the opposite edge. Let ABC be a section by a plane perpendicular to OA. Draw AP perpendicular to BC. Then AOP is the pyramidal error, p, as above defined. $B\hat{A}C$ is the angle between the planes AOB and AOC $=\hat{A}$ say; but angle ABC is not the angle between the planes AOB and BOC, because AB and BC are not perpendicular to OB. If the angle between the planes is $\hat{B}$, $\hat{B} > A\hat{B}C$. Similarly $\hat{C} > A\hat{C}B$. Thus $\hat{A}+\hat{B}+\hat{C} > \hat{A}+A\hat{B}C+A\hat{C}B$. These are the angles of a plane triangle, $\therefore\ \hat{A}+\hat{B}+\hat{C} > 180°$. Denote $O\hat{B}A$, $O\hat{B}C$, and $A\hat{B}C$ by α, β, and γ respectively; then from the geometry of the solid angle of which α, β, γ, and $\hat{B}$ are elements,

$$\cos\gamma = \cos\alpha\cos\beta + \sin\alpha\sin\beta\cos\hat{B},$$

$$\therefore\ \cos\hat{B} = \cos\gamma\operatorname{cosec}\alpha\operatorname{cosec}\beta - \cot\alpha\cot\beta,$$

$$= \cos\gamma\,(1+\tfrac{1}{2}\cot^2\alpha+\tfrac{1}{2}\cot^2\beta) - \cot\alpha\cot\beta,$$

since α and β are nearly right angles.

$$\cot\alpha = \frac{AB}{AO} = \frac{AB}{AP}\cdot\frac{AP}{AO} = \frac{p}{\sin\gamma};$$

similarly $$\cot\beta = \frac{p}{\tan\gamma}.$$

$$\therefore\ \cos\hat{B} = \cos\gamma\left\{1+\frac{p^2}{2}\left(\frac{1}{\sin^2\gamma}+\frac{1}{\tan^2\gamma}\right)\right\} - \frac{p^2}{\sin\gamma\tan\gamma}$$

$$= \cos\gamma - \frac{p^2}{2}\cos\gamma,$$

$$= \cos\left(\gamma + \frac{p^2}{2}\cot\gamma\right),$$

$$\therefore\ \hat{B} = \gamma + \frac{p^2}{2}\cot\gamma.$$

For an isosceles prism, it is clear from symmetry that $\hat{C} - A\hat{C}B$ is equal to $\hat{B} - \gamma$. Therefore ξ, the excess of the three prism angles over 180°, $=2(\hat{B}-\gamma) = p^2\cot\gamma$.

In *Fig.* 10 the values of ξ for various pyramidal errors are plotted for the two

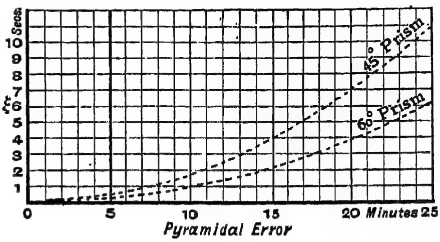


FIG. 10.

commonest prisms. We see that the sum of the angles may be taken as 180° without appreciable error if the value of p does not exceed 3 or 4 minutes. For larger values ξ increases rapidly and has to be taken into account in precise measurements. We may measure p on the spectrometer as follows. Let ABC, *Fig.* 11, be the prism, arranged as

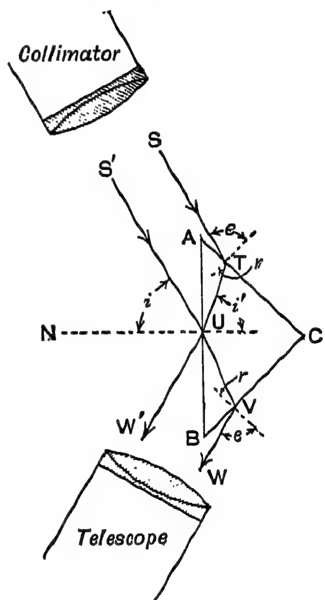


FIG. 11.

shown with respect to the collimator. Rays such as S′U are reflected from the face AB in the direction UW′. Rays such as ST follow the path STUVW, VW being parallel to UW′ if $\hat{A}=\hat{B}$ and $p=0$. In the presence of pyramidal error one will slope upwards and

the other downwards. Their vertical displacements can be measured with the micrometer eyepiece suitably oriented. Take as reference plane a plane perpendicular to the edge C, and let the incident light be in this plane. The normal NU will have an inclination p to the plane. Suppose it slopes downwards towards the left. Then the reflected ray UW′ slopes downwards by an angle $\phi=2p\cos i$. Since AC is perpendicular to the reference plane TU is also in this plane, so the ray UV slopes upwards by an angle $\phi'=2p\cos i'$. This ray meets BC, which is also perpendicular to the reference plane, and is refracted from glass to air. It is easily shown that ψ, the upward inclination of the emergent ray VW, is equal to $\mu\phi'$, where μ is the refractive index of the prism.

Thus the vertical displacement, δ, of VW with respect to UW′

$$=\phi+\psi=2p\,(\cos i+\mu\cos i').$$

We must substitute for i' in terms of the angles of the prism and the angle of incidence i.

$$i'=\hat{A}+r=\hat{A}+\sin^{-1}\frac{1}{\mu}\sin e,$$

$$=\hat{A}+\sin^{-1}\frac{1}{\mu}\sin\{\pi-(\hat{A}+i)\},$$

therefore

$$\delta=2p\left[\cos i+\mu\cos\left\{\hat{A}+\sin^{-1}\frac{1}{\mu}\sin(\pi-\hat{A}-i)\right\}\right],$$

whence p is obtained.

The advantage of this method is that δ is not appreciably affected by incorrect adjustment of the prism with respect to the plane of incidence; the effect is simply to introduce a cosine term. In carrying out the measurement, the two images of the slit will be superposed if $\hat{A}=\hat{B}$. To observe δ it is necessary to block out first the rays externally reflected and then those internally reflected.

Pyramidal error may also be measured by auto-collimating methods. In *Fig.* 7 (*b*), for instance, it is obvious that since the γ edges of the prism OPQ and its reflection OP′Q′ are coincident the faces PQ and P′Q′ have a *vertical* inclination of $2p$. Thus the vertical component of the displacement of HI″ with respect to AI′ $=4\mu p$. Similarly in *Fig.* 8 (*b*) the $\beta\gamma$ plane of the reflected prism is parallel to the β edge of the actual prism and is therefore inclined vertically with respect to the $\alpha\gamma$ face of the latter by p. Hence in this case the displacement of the two reflected beams has a vertical component $2\mu p$.

J. G.

Grainless Photography, use for graticule production. See "Graticules."

Gramophone. See "Phonograph and Gramophone."

GRATICULES

Graticules, a term which came into common use during the Great War, may be defined as being the measuring marks or scales, usually on a glass plate or disc, placed in a focal plane of an optical instrument for determining the size, distance, direction, position, or number of the objects viewed coincidently with the scale itself. The term "graticule" comprises, therefore, the numerous appliances variously known as sighting scales, reticules, cross-lines, eyepiece, and stage micrometers, diaphragms, webs, etc., which are used in telescopes, microscopes, and other optical instruments, and it is usual to refer to the disc or plate with the marks on it as a "graticule."

Being normally viewed under considerable magnification, it is of primary importance that the lines or marks should be quite sharp and clean, whatever the thickness of the lines may be. The thickness may vary between ·0001″ and ·004″ according to the purpose for which they are used and the method of manufacture. The lines may be required to be opaque, or transparent, or to have special optical characteristics, and, except in the case of simple cross-lines, the divisions or spacing of the graticules usually demands the very highest degree of accuracy. Further, where the graticule takes the form of a circular disc inserted within an optical instrument, very exact centring of the pattern on the disc is an essential. Owing to these exacting demands in the way of accuracy and precision, the manufacture of graticules is not the simple matter which it would appear to be, and has been the subject of an immense amount of research.

The methods of manufacture may be classed under two main headings: (*a*) mechanical ruling, (*b*) photographic methods. The methods by mechanical ruling may be subdivided into—

(1) Ruling with a diamond on glass.

(2) Ruling with a diamond, and subsequently filling in the marks with graphite, or some other opaque substance, the latter being usually mixed for the purpose with copal varnish or some similar medium.

(3) Coating the glass with a transparent film of collodion or similar medium, and ruling the lines on this, subsequently cementing on a cover glass.

(4) Coating the glass with waxy or bituminous substances, ruling through this, then etching the lines by means of hydrofluoric acid gas, or by immersion in solutions containing hydrofluoric acid, and, lastly, filling in with opaque substances.

The photographic methods may be subdivided into—

(1) Photographs on plates or discs with

albumen or collodion emulsion having a grain as fine as possible.

(2) Photo-ceramic methods.

(3) Photo-etching methods.

(4) By Grainless Photography.

(5) By Filmless Photography.

Of these various methods Nos. 1 and 4 of the photographic methods require a cover glass cemented on to protect the film; all the others are " coverless " graticules.

In all countries the principal method of manufacture up to the year 1914 was the mechanical method, the marks being usually filled in with an opaque substance. Simple patterns, such as straight lines at equidistant intervals or cross-lines, can be produced very perfectly in this way, but the method becomes cumbersome when anything more elaborate is required, such as scales with varying line thickness, or curved lines or numbered lines. Further, where lines cross each other the glass is apt to be broken up, causing the critical point to be less good than the rest, and this has led to the practice, when ruling lines at right angles to each other, of stopping the line just short of the cross-line, and continuing it again a short distance from the other side. The sharpness and apparent opacity of diamond-ruled lines depend on the cleanness and fineness of the diamond cut and personal skill of the operator; the diamond points are apt to wear out fairly rapidly and then produce inferior work.

Ruling and etching methods are not so subject to wear and tear of the ruling point, because the waxy or bituminous substance ruled through is soft. Curved lines, numerals, or letters can be reproduced by a pantograph arrangement. In other respects they suffer from the same limitations in manufacture as diamond-ruled graticules, and it is exceedingly difficult to coat them with a resist so perfect that the hydrofluoric acid does not attack the glass here and there through the resist, occasioning tiny marks or spots on the glass which are objectionable.

Graticules produced by any method of mechanical ruling require separate testing for accuracy and finish, each being an individual piece of work standing on its own merits.

Graticules produced by photographic means have a great advantage in this respect, for any number of graticules can be produced exactly similar to one another, without requiring separate testing for accuracy of scales, always provided that the photographic method employed is one in which the film is given no chance of expanding or contracting during the process. The main security for this is by only employing such processes in which the film is not transferred from one glass to another, *i.e.* each graticule disc or plate must be coated separately and the film must be left *in situ* throughout the operations.

Many fine grain processes have been employed, collodion in general having been found to be the most suitable medium, since gelatine coatings almost invariably accumulate dust specks and are not sufficiently clear. Carbon processes on bichromated gelatine or fish glue have likewise been employed; in this case all except the image lines, which are rendered insoluble, is removed by washing. Such processes yield very fair results in the case of graticules with comparatively coarse lines, or which are subject only to low magnifications.

Photo-ceramic methods have been employed in Germany mainly for one particular purpose, viz. graticules for binoculars. These graticules have a transparent image, which is more or less faint—an advantage for the purpose in question. They show little evidence of grain. The process has been kept secret, but some ideas as to the methods by which they are produced may be inferred from German and Austrian photo-ceramic literature.[1]

Messrs. Adam Hilger, Ltd., of London, have also developed a photo-ceramic method. These graticules have a fine grain, the image lines being sufficiently firmly adherent to the glass surface to stand the necessary cleaning of the discs. They will also withstand ordinary solvents. This process would seem to have been employed mainly for graticules with fairly coarse lines, used with low magnification.

Photo-etching methods which were worked out by J. Rheinberg, London, in the early years of the War, by coating discs with an emulsion impervious to the action of hydrofluoric acid gas before exposure to light, but pervious to its action after exposure, have not hitherto proved successful, because the resist suffered from the same defects already referred to as liable to occur in the etching methods after mechanical ruling.

Methods of graticule production by photography on a sensitive collodion film, having a grain so fine as to stand considerable magnification, date back a considerable time, and until recent years were chiefly practised in Germany. The actual process has been kept secret, but there is evidence to show that it has followed on the lines of the old Taupenot Collodion Process, dating from 1855, or some modification of the process.[2]

Messrs. Rheinberg & Co., London, have in recent years introduced graticules made by a secret process styled " Grainless Photography," which differs widely from any

[1] *Die Photokeramik* by Julius Krüger and Jakob Husnik, published by A. Hartleben, Vienna and Leipzig.

[2] For an account of these processes readers are referred to Hardwick's *Photographic Chemistry*, 1859, and an article on "Fine Grain Photographic Plates," *British Journal of Photography*, September 13, 1912.

previous process, both as regards the composition of the film, the chemicals employed, and the methods of photographic treatment. Collodion is used as the vehicle of the image. By different modifications of the process, either black opaque lines are produced, or white opaque lines, or transparent lines in grey, amber, violet, or other tints. The lines remain clean and sharp under considerable magnification, the transparent ones show no vestige of grain under high magnifications.

Graticules produced by so-called "Filmless Photography," another secret process developed by the brothers Rheinberg during the Great War, represent the latest advance. These are similar in appearance and characteristics to those by "Grainless Photography," and may be opaque or transparent, but the image vehicle is the surface layer of the glass itself, in which the lines are formed in untarnishable metal. In a perfect filmless photograph the lines cannot be rubbed or cleaned off, and they are unaffected by solvents save one or two powerful acids, chiefly those which attack the glass itself.

In one special modification of this process, termed the "Silverline" process, a brilliant white image consisting of an extremely fine fibrillar interlaced network of granulated silver, is produced upon the glass surface, and not incorporated within the surface layer. This image, unlike that in any other kind of graticule, has the property of reflecting an approximately equal amount of light in all directions, no matter at what angle the light is incident upon it. The image, however, requires protection by means of a varnish.

In a paper on "Graticules" (*Transactions of the Optical Society*, May 1919) Mr. Julius Rheinberg has stated the general methods and considerations relating to their photographic production (with special reference to those made by Grainless and Filmless Photography), together with considerations as to their design in relation to the various purposes for which they are required, and in the main the following particulars are abstracted herefrom.

The usual way is to start with a greatly enlarged drawing, say from 12″ to 15″ diameter, from which an intermediate reduced negative is produced, and this is again reduced in a precision camera to the exact size. The positive so obtained serves for the production of the "Master negatives," from which the graticules are then made by "contact printing." Drawings should be on smooth, thick, Bristol board, not on thin paper or linen which has a tendency to cockle. For fine work the nature of the ink is important, and of a number of Indian inks tried, Messrs. Winsor & Newton's "Mandarin Black" was found to be the best.

Whilst it is not needful to draw the original to any particular scale of magnification, it is very needful to bear in mind that all photographic processes increase the relative thickness of lines, partly owing to the grain of the plate, partly owing to optical considerations, and, to compensate for this, the lines of the drawing should be finer than the actual scale of magnification of the drawing. Even in graticules produced by Grainless and Filmless Photography the intermediate negative has to be made on a fine *grain* plate, because grainless emulsions are too slow to admit of camera reductions from drawings on an opaque board, and lines intended ultimately to be—

1/500″ should be drawn about 5 per cent thinner than the scale magnification.
1/1000″ should be drawn about 10 per cent thinner than the scale magnification.
1/2500″ should be drawn about 50 per cent thinner than the scale magnification.

The optical considerations referred to above relate to the limits of accuracy germane to all reproductions by photographic lenses. No lens produces an equally sharp image on a flat plate over the whole field, but the deviations are small enough to be negligible with a good lens in ordinary photography because the lines are much coarser and the grain of the plate itself is usually the determining factor in setting the limit. When, however, lines of ·0005 inch and less are in question, the lens sets the limit, as it is difficult to secure these evenly over more than a comparatively narrow angular field.

Similarly the lenses themselves set a limit to the accuracy with which long scales can be reproduced by photographic reproduction, as slight errors may occur in the spacing of divisions owing to zonal differences of magnification.

For these reasons many kinds of graticules cannot be properly produced from large drawings by camera reduction at all, the "master negatives" being made by contact printing from originals mechanically ruled. Where figures, etc., are employed combinations of photographically reduced drawings and mechanically ruled scales can be used to produce the master negatives.

In this manner grainless photographs have been produced with lines as fine as 1/12500″, although lines finer than 1/5000″ are a matter of considerable difficulty.

In any photographic process of graticule production the order of accuracy between one scale and another, when large quantities of a scale are produced, is a matter of importance. This will depend entirely on the nature of the process, also whether any film transfer has taken place. In the case of Grainless and Filmless Photography a number of scales were sent to the National Physical Laboratory, taken from batches at random, which had all been produced from the same master negative, but not actually from the same negatives, as a master negative serves for making any number of the actual negatives used by means of intermediate positives, and the tests showed the maximum error of the scales, either as regards over-all length, or as regards the divisions, to be within one-twentieth per cent, and consequently since this error included any

actual error in the original master negative, the actual variations between one scale and another produced by either of the processes in question would be below that amount. The main test was made on scales of 10 mm. length; tests on smaller numbers of scales up to 100 mm. length gave similar results.

The vehicle bearing the image is an important consideration in graticules produced photographically. Collodion has in general been found to be the most suitable and is employed in the Grainless Photographic process as well as in the foreign-made graticules.

In Filmless Photography, however, the vehicle which bears the image is the surface layer of the glass itself, and the nature and composition of the glass play an important part, since the process is based on the idea that as glass is porous, though the pores may be ultra-microscopic, matter in a suitably fine state of division can be introduced into these pores by appropriate methods. The most suitable kinds of glass have been found to be the various descriptions of crown and plate glass. Glasses of a particularly dense nature, such as barium glasses and baryta light flints, have been found less suitable, although with more or less trouble filmless photographs can be produced thereon, particularly transparent photographs which do not necessitate the introduction of as much metal into the surface layer of the glass as opaque ones. Glasses containing any considerable proportion of lead, *i.e.* the ordinary description of flint glasses, are unsuited to the process, though this is due to the softness of these glasses and chemical combinations which take place rather than to any question of density.

The uses to which graticules are put are fairly numerous; they form a most essential part or adjunct of the optical instrument in or with which they are employed, and there are but few optical instruments used in peace or war in which they are not either permanently or temporarily used. A good example of the relative importance the graticule may assume in an optical instrument is afforded by the Aldis Unit Telescopic Sight for aeroplanes, the whole instrument being merely a device for seeing the graticule image sharply, coincidently with the target which is viewed without any magnification whatsoever.

The kind or variety of the graticule naturally depends upon the purpose for which it is used in an optical instrument, and the advent of Grainless and Filmless Photography which enables graticules to be produced with image lines opaque, transparent, or semi-transparent, in various colours or tones and possessing different optical properties, has not only considerably increased the purposes for which they can be employed, but also in some instances given rise to new forms of instrument construction. The majority of graticules are required for use with transmitted light, and for most of these it has been customary to use graticules having opaque black lines, the possible advantage in some cases of transparent lines not having received any adequate attention, probably owing to the fact that such graticules were not generally available.

The optical difference between a neutral coloured transparent line and an opaque one is that the former reduces light from the object by a definite percentage, whereas the latter cuts out the whole of the light. There would appear to be a prevailing impression that transparent lines will not stand out sufficiently, or show sufficient contrast against a dark ground, which, however, experiment will show not to be based on fact. Excessive contrast induces eye fatigue and loss of comfort in use, which becomes particularly noticeable and indeed irritating when graticules have opaque patterns in the nature of squared lines or other repeating patterns. Such contrast may hinder the eye in picking up and inspecting critically the object to be viewed. An outstanding merit of transparent lines is that they do not cut off any part of the object viewed, a not unimportant matter where lines are fairly coarse, *i.e.* exceeding ·001 inch, which are frequently required. Other advantages of transparent lines are referred to later on in the notes on graticule design.

For one class of graticules, viz. military binoculars, transparent lined graticules have, however, been in use for some considerable time; for this purpose the German photo-ceramic graticules were in use prior to the war, also etched or directly ruled lines on glass, not filled in. They afford a typical example of the benefit of faint transparent lines, as they do not distract the eye from the object and are little noticed, unless attention is purposely concentrated on them.

Another class of graticules, used typically in sighting instruments for day and night use, are intended for use both with transmitted light and reflected light, appearing black when illuminated in the former way, and white in the latter. They are produced by filling in a diamond-ruled image with white opaque lines, or even more satisfactorily by a modification of the grainless photographic process, which yields white opaque lines. Their opacity renders them black by transmitted light, and for night work the light is introduced through the edges of the disc by a subsidiary lamp, the light so introduced being imprisoned in the glass except where it strikes the image lines, which reflect a portion of it towards the eye lens.

This class of graticules would seem susceptible of development to numerous uses, for example spectroscope eyepieces, as, by illuminating the disc

edge with coloured light, the graticule image may be made to approximate in colour to the part of the spectrum under observation. It is well known that the eye will not sharply focus widely differing colours, or colours and white lines simultaneously, hence the utility of some such device.

Graticules for use by reflecting light only are in general produced by the same processes already referred to. They are usually illuminated either directly or by reflectors prepared with white-diffusing substances.

Whilst in all ordinary cases a graticule is so mounted that the plane of the graticule disc is parallel with the lenses of the optical system with which it is used, an exception occurs in certain military instruments, in which the graticule has to be viewed through the optical system at continuously varying obliquities, whilst the direction of the light

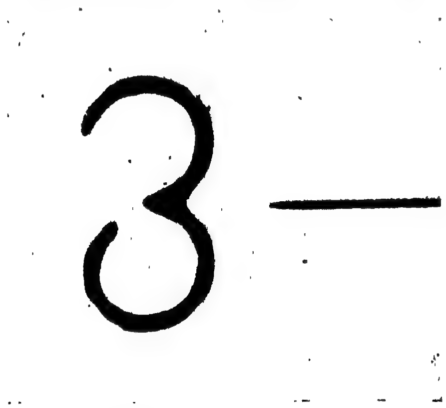


FIG. 1.—Black opaque. Line thickness ·0004″. Magnification 85.

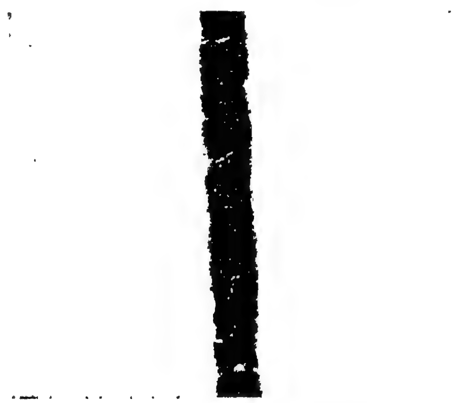


FIG. 2.—Black opaque. Line thickness ·0004″. Magnification 650.

which is reflected from the graticule lines remains more or less constant. Ordinarily the intensity of the light reflected by a line or a surface varies greatly according to the angle at which the light impinges on it, as a consequence of which the image of the graticule would be far less brilliant when viewed at certain degrees of obliquity than at others. To meet this case the "Silverline" process already referred to was devised, in which the image lines have the property of reflecting or diffusing light of approximately equal intensity in all directions, irrespective of the angle at which the light impinges on them.

Another class of graticules are such as have clear lines on a black opaque ground. These are illuminated from behind or by a separate source of artificial light. They are mostly used in the type of instrument in which it is desirable to project graticule lines, etc., in the image plane, but in which it is not convenient to place the graticule in the main optical system, either because that system has no real focal plane, or for other reasons of design. A frequent device is to place the graticule in the focus of a collimating lens and place a semi-transparent reflector in the main optical system at an angle of 45°. Graticules of this description are produced by Grainless Photography, and are similar to the negatives ordinarily employed in that process.

The design of graticules is a matter repaying careful attention. It should not be overlooked that the graticule is often an essential, and frequently a vital part of the optical instrument, that it is under constant scrutiny, and should be so designed as to secure the maximum efficiency and the greatest amount of comfort in use. To ensure this it is absolutely necessary that it should be properly correlated to the conditions under which it is used and to the general optical system of the instrument.

It is desirable to consider—

(1) The nature of the pattern.
(2) The thickness of the lines.
(3) Questions of opacity or transparency.
(4) Questions of differentiation or colour.
(5) Special optical desiderata.
(6) Questions affecting price or difficulties in production.

Regarding the nature of the pattern, an important point is to keep it as simple as possible. Every superfluous line, figure, or other mark should be eliminated, for these only occupy the field otherwise wanted for examination of the object and distract the eye. The fact that additional lines, figures, and so forth do not add commensurately to the cost of graticules produced by photographic processes is a temptation to load them up unnecessarily, but should be resisted, as

practical experience during the War has frequently led to complicated patterns being successively simplified.

With regard to the line thickness of graticules, a number of interesting little points require to be borne in mind and balanced, and in considering these it is convenient to remember that a line 1/1000″ wide, viewed with the unaided eye at normal, *i.e.* 10″ distance, subtends an angle of about 20 seconds. Opaque lines when viewed against a white ground appear rather thinner than they actually are, as the white light encroaches on the two edges of the line, by irradiation of the eye. When the opaque line is too narrow, the line appears faint and indistinct and begins to be uncomfortable. With the unaided eye a line 1/500″ or 1/20 mm. is sufficiently narrow for comfort. For eyepiece graticules the magnification of the eye lens has to be taken into account, and the best rule for comfort is not to have lines narrower than needful. With eyepieces magnifying × 5 to × 8 opaque lines 1/2000″ or 1/80 mm. will be found sufficiently fine. In micrometer scales divided into tenths or twentieths of a millimetre the width of the lines may reasonably be reduced to 1/5000″ or 1/200 mm., and in scales of this description, or scales divided even more finely for microscope stage micrometers of the ordinary description, opaque lines are the most suitable. The illustrations, Figs. 1 to 4, are from photomicrographs of Rheinberg Filmless Graticules taken by Mr. J. E. Barnard, F.Inst.P.

Transparent lines seen against a white ground always appear markedly narrower than opaque ones of similar width, due partly to the greater encroachment on their edges by irradiation. It is rarely necessary or advisable to go below ·001″ in width, the more so as they have the advantage that the object viewed can be seen through them, and centred to the middle of the transparent line with very great accuracy. This feature is also useful because lines of varying thickness can be used on the same graticule disc for differentiation purposes, without sacrificing accuracy of measurements.

Where squared lines or repeating patterns, or markings in the nature of massed patches, are required, transparent lines should always be given preference over opaque ones, unless the latter are indispensable for reasons necessitating *maximum* contrast. The depth of tint of transparent lines should be appropriate to the magnification under which they are to be viewed, as it is a somewhat remarkable fact that it differs with the magnification. The same grey line of ·002″ width may look almost black to the unaided eye; seen with an ocular × 5 it appears grey, with an ocular × 10 a lighter grey, on the microscope stage magnified 100 times it appears a very light grey indeed.

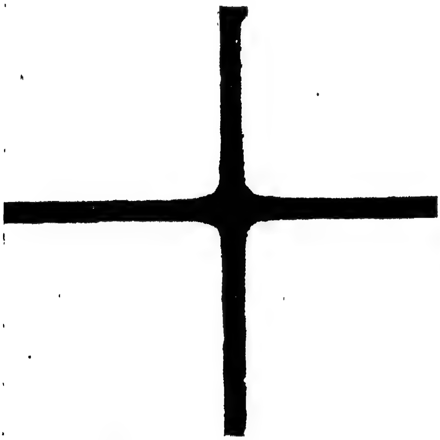


FIG. 3.—Black opaque. Line thickness ·001″. Magnification 85.

FIG. 4.—Black opaque. Line thickness ·001″. Magnification 650.

Differentiation for the sake of clearness and facility in reading or counting is a factor which should receive consideration. For instance, in a graticule for counting purposes, a field consisting of alternate transparent grey and colourless square patches in a chessboard pattern—the so-called Chessboard Micrometers—are more comfortable and easy to work with than graticules simply divided into squared lines. Lines of different thickness on the same graticule are another form of differentiation. Cases occur, however, where it is still more convenient to differentiate by lines of different colour on the same graticule. Such graticules

were patented by Messrs. Aldis Bros., Birmingham, and known by name of composite graticules, because they are usually produced by doing the different colours by Grainless Photography on two separate discs and cementing them together. The colours themselves, in so far as it is possible to produce them, may be selected to obtain contrast in accordance with the purpose for which they are used; for example, an aeroplane telescope graticule for use against the blue sky might have golden amber lines, one principally for use against green foliage would conveniently have lines of a violet tone.

When one scale is measured against another scale (microscope stage micrometers and eyepiece micrometers offer a familiar example) there is an advantage in having the scales differentiated in colour so that they can be distinguished at a glance.

The special optical desiderata which the graticule should possess will depend chiefly upon the methods of illumination under which it is to be used, *i.e.* whether by transmitted or reflected light or edge illumination, whether it is to be viewed normally to its surface or otherwise, whether with colour screens or such-like considerations. These matters have already been referred to.

Lastly, a graticule designer should bear in mind that the cost of production of any pattern frequently depends largely on seemingly quite trivial matters, and should consult with the manufacturers in order to avoid specifications which unduly increase cost without increasing efficiency. For example, specifying a graticule on a disc 10 mm. diameter, when a disc of 15 mm. diameter might be adopted just as well, might entail a difference of 50 per cent in cost of production. Or by requiring the cross-lines in a Filmless Photograph to go right up to the edge of the disc, a large extra cost may be incurred compared with one in which the lines leave off ½ mm. or 1 mm. from the edge of the disc. Very fine lines in immediate contiguity to very coarse ones are inadvisable, as they require different treatment by all photographic processes. In short, the instrument-maker should remember that the graticule, which is at least as important as the other parts of his optical instrument, should not be treated as an independent adjunct, to be fixed up after everything else is settled, but should be carefully considered along with the design of the whole instrument, if he is anxious to obtain maximum efficiency at lowest cost.

Consequent upon the limitation of patterns owing to the restricted methods of production available until recent years, the application of graticules has not been developed in many directions in which they would appear susceptible of extension. During the Great War the tendency manifested itself for the design of new types of optical apparatus which would not have been feasible but for the fact of the facility with which new and unusual types of graticules could be produced, and it seems likely that many types of optical apparatus might be improved, or new types designed, by paying due regard to the possibilities which graticules now afford. It is only needful to bear in mind, for instance, that the steel slits of spectroscopes could be replaced by graticules having a series of perfectly sharp, clean, colourless lines on a black ground of definite widths, and that as the films have less thickness than steel jaws there is less liability to stray reflections or diffraction. Or, again, the fact that completely divided circles can be produced with great accuracy in the form of a small graticule opens up the possibility of designing instruments in which such circles are either placed in, or projected into, the plane of optical instruments to supplant the large and expensive graduated metal arcs and circles with their separate reading-microscopes elsewhere. Such examples show that the possible applications of graticules with a view to improvements or simplification in the design of optical instruments is a subject which will well repay study, and has indeed its own place in the general study of Applied Optics.

J. R.

Graticules, Stereoscopic: a device employed in some stereoscopic rangefinders to give a distance scale in the field of view. See "Rangefinder, Short-base," § (13).

Grease-spot Photometer: another name for the Bunsen photometer. See "Photometry and Illumination," § (15).

Grinding and Polishing of a Glass Surface, Stages in the. See "Optical Parts, The Working of," § (7).

Groove Form: effect on distribution of light in diffraction spectra. See "Diffraction Gratings, Theory of," § (8).

Guild Spherometer. See "Spherometry," § (5).

Guitar: a musical instrument of six strings which are plucked by the right hand. See "Sound," § (26).

Gyro, Floating Ballistic. See "Navigation and Navigational Instruments," §§ (13) (ii.), and (14).

Gyro Compass. See "Navigation and Navigational Instruments," §§ (13), (14), (15).

Gyro Compass, Speed Error of. See "Navigation and Navigational Instruments," § (15).

γ-Rays, General Properties of. See "Radioactivity," § (14) (i.).

Nature of. See *ibid.* § (14) (iv.).

— H —

HALF-SHADOW ANGLE, VARIATION OF SENSITIVENESS OF POLARIMETER WITH. See "Polarimetry," § (7).

HALF-VALUE PERIOD OF A RADIOACTIVE SUBSTANCE: a term used in radioactivity to denote the time taken for the atoms present in a radio-element to decrease to half. See "Radioactivity," § (6).

HARCOURT PHOTOMETER. See "Photometry and Illumination," § (26).

HARDNESS OF GLASS. See "Glass," § (25).

HARDNESS OF X-RAYS: quality of X-rays. See "Radiology," § (17).

HARP: the most important representative of stringed instruments which are played by plucking with the fingers. See "Sound," § (28).

HARRISON STREET PHOTOMETER. See "Photometry and Illumination," § (57).

HARTMANN'S METHOD OF DETERMINING THE POSITION OF AN IMAGE-PLANE. See "Objectives, Testing of Compound," § (1).

HAÜY'S LAW OF CONSTANCY OF CRYSTAL ANGLES, final proof of, from the general law of progression of crystal properties corresponding to the advance in atomic number. See "Crystallography," § (13).

HEARING, the act of. See "Sound," § (57) (ii.). Human. See *ibid.* § (57).

HEELING ERROR in magnetic compass. See "Navigation and Navigational Instruments," § (10) (ii.).

HEFNER LAMP: a flame standard officially adopted in Germany. See "Photometry and Illumination," § (7).

HEIGHT, DEFINITION OF. See "Surveying and Surveying Instruments," § (36).

HEIGHT-FINDER, THE PATERSON-WALSH ELECTRICAL

§ (1) THE BENNETT-PLEYDELL PRINCIPLE.—For determining the height of an object, as distinct from its range, the simplest geometrical arrangement is that involving the measurement of the distance between two stations and of two angles. This was suggested in connection with the determination of the height of aircraft by Lieutenant Mansell Pleydell and Dr. G. F. Bennett, and is known as the Bennett-Pleydell principle.

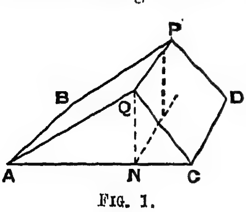


FIG. 1.

In *Fig.* 1, if two planes ABPQ and CDPQ be rotated about their axes AB, CD, until they intersect in PQ, then if AB and CD be parallel to each other, perpendicular to the line AC, and in the same horizontal plane, every point on the line PQ will be at the same height above this plane, this height being given by the formula $h = AC/(\cot A - \cot C)$, where A and C are the angles (both measured in the same sense) which the planes ABPQ and CDPQ make with the horizontal.

For in *Fig.* 1 let AQC be a vertical plane through the base AC, and QN perpendicular on AC. Then

$$QN = h, \text{ also } QAN = A, \ QCN = 180^\circ - C.$$
$$AN = QN \cot QAN = h \cot A.$$
$$CN = QN \cot QCN = -h \cot C.$$
$$\therefore AC = AN + CN = h(\cot A - \cot C).$$

§ (2) THE TELEPHONIC INSTRUMENT.—The height-finders which have been designed on this principle are long-base instruments. The first (telephonic) type consists of two stations, one at each end of a base about a mile long. At each station there is a wooden stand (*Fig.* 2) holding a rectangular sighting frame F,

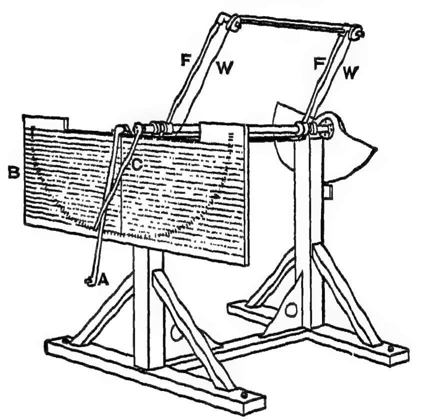


FIG. 2.

capable of rotation about a horizontal axis. At one end of each instrument is a dial and pointer arrangement giving at once the angle between the horizontal plane and the plane of the sighting wire WW attached to the rotating framework F. At the other end of *one* instrument is a "plotting-board" B, rigidly attached to the upright and engraved with a series of equidistant horizontal lines marked in thousands of feet. In addition to these lines there is a semicircular scale of degrees over which a loose arm A may be swung at will. Rigidly attached to the rotating framework F is a second arm C. The inner edge of each of these arms passes

through its own axis of rotation, and the distance between these axes represents the distance between the "home" and "distant" instruments on the same scale as the vertical distance between either axis, and any one of the horizontal lines represents the height marked on that line.

In using the instrument the sighting frames at both stations are set with their axes parallel and at right angles to the base and are turned so that the plane of each wire rectangle passes through the object. The angle of elevation of the "distant" plane is then telephoned to the "home" station, and the arm A is set to this angle on the plotting-board. The arm C automatically gives the same information for the "home" station; the triangle formed by the inner edges of the arms A and B, with the horizontal line passing through their axes, is an inverted scale reproduction of the triangle AQC of *Fig.* 1. Hence the height of the object sighted may be read off from the position of the intersection point of A and B.

§ (3) The Electrical Instrument.—The chief disadvantage of the telephonic instrument is that its readings, being dependent on communication of the angle of elevation at the distant station, are necessarily discontinuous and possess a certain small time lag. To obtain a continuous indication of height the electrical instrument was devised. In this the use of the plotting-board and scale of angles was dispensed with except as incidental to the periodical checking of the electrical adjustments. At each station there is a uniform resistance supported horizontally at a fixed distance below the axis of the sighting frame, and with its centre directly under this axis. Attached rigidly to the axle of this frame, and in continuation of the plane of the sighting wires, is a phosphor-bronze bar which presses against the horizontal resistance. In this way the distance between the *centre* of the resistance and the point of contact of the bar is always proportional to the cotangent of the angle of elevation of the sighting frame (see *Fig.* 3). Thus if it

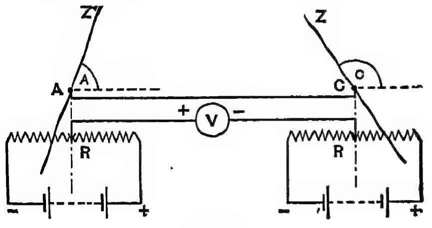


Fig. 3.

be arranged that there is a constant current passing through the resistance, the difference of potential between the bar and the centre of the resistance is also proportional to this quantity

By means of a local battery, voltmeter, and regulating rheostat, the voltage drop across a given length of resistance is maintained at the same constant value at both home and distant stations. The distances between the axes of rotation of the sighting frame and the horizontal tracks of the phosporbronze bars on the resistances are also the same at both stations, so that by connecting line conductors between the stations in such a manner as to put the two voltage differences in series, a milliammeter with a suitable swamping resistance in the circuit gives at once a reading proportional to (cot A − cot C). It follows that, in order to obtain a direct reading of height, the instrument must be provided with a reciprocal scale. For the purpose of increasing the range two scales are provided, one from 1600 to 5000 ft., and the other from 5000 to 20,000 ft. The more open scale is brought into operation by pressing a key switch which normally short-circuits a shunt to the moving coil of the instrument.

§ (4) The Paterson-Walsh Instrument.—The series resistance of the milliammeter is variable, and is so adjusted in relation to (i.) the actual voltages employed at the two stations, (ii.) the sensitivity of the milliammeter, and (iii.) the ratio of the base length to the vertical distance between the axes of the sighting frames and their respective resistances, that a given base length corresponds with a definite series resistance for all instruments. Corresponding values are tabulated in the instrument box, so that any instrument can be used with any length of base by suitably adjusting the series resistance. The electrical circuit used in the Paterson-Walsh design of the instrument is shown diagrammatically in *Fig.* 3. R, R are the uniform resistances, AZ, CZ the sighting planes, and V the milliammeter. In this diagram the local arrangements for maintaining a constant and definite current through the resistances are not shown.

The chief errors to which the instrument is liable are (i.) leakage in the insulation of the main conductors between the stations, and (ii.) change of resistance of these conductors with change of temperature. On this latter account these conductors are required to be as heavy as possible, so that their resistance may be properly swamped by the series resistance of *V*.

§ (5) Other Instruments.—It will be clear that other adaptations of the principle of the electrical solution could readily be devised. One of these avoids the use of a battery at each station. The two portions of the regular resistances are put in series and form one arm of a Wheatstone bridge. By continuously adjusting a local resistance

to keep the bridge in balance, a measurement of the height may be obtained from the value of this resistance in circuit at any time.

The accuracy of all height-finders working on the Bennett-Pleydell principle is necessarily dependent on the parallelism of the axes of the two sighting frames, and it is desirable that the departure from this parallelism should not exceed 6′ of arc. If parallel, the axes may be slewed with respect to the base line by as much as 5° without correction, or by any angle up to 30° if the length of the base line be corrected by the factor cos θ.

§ (6) Correction for Difference of Level.—In all that has been said above it has been assumed that the two stations are on the same level. If their difference of level be d, it may be readily shown that the new formula for h is

$$\frac{\{b-(d/2)(\cot A+\cot C)\}}{(\cot A-\cot C)}.$$

A rigorous practical solution of this expression is not possible, but a very close approximation, in all practical cases, to the correct result is obtained by setting over the sighting frames with respect to their pointers, so that when the latter are reading 90° the frames are not vertical, but are tilted downhill *at both ends* by an angle equal to half the angle of the slope (supposed continuous) between the two stations. In this manner the instrument indicates the values of $b\,(\cot \overline{A-\theta}-\cot \overline{B-\theta})$, where $\tan 2\theta = d/b$, and it may be shown that this formula gives a very close approximation to the accurate expression, heights being indicated as above the *mean level* of the two stations.

§ (7) Use of the Instrument.—The instrument can be used for all angles of elevation greater than 20°, but below this value the line of contact of the phosphor-bronze bar on the even resistance is liable to variation, and a small inaccuracy of sighting produces a considerable error in the indicated height. The error of an instrument such as that described should not exceed 300 ft. at a reading of 12,000.

It will be apparent that if sight of the object be lost at one station an entirely erroneous reading may be given by the indicating instrument if no means is provided for cutting it out of circuit. In the actual design, the circuit connecting the two stations with the indicator is not completed until a small key switch is depressed at each station. These keys are operated by the observers at the sighting frames, so that if the target be lost at either end, the indicator can be at once brought to zero, *i.e.* to infinity height on the reciprocal scale.

J. W. T. W.

Height-finders, Anti-aircraft. See "Range-finder, Short-base," § (8).

Heliostat, The: a clockwork device for directing the rays from the sun into a fixed telescope or other optical system. See "Telescope," § (17).

Heterochromatic Photometry: the comparison of lights of different colours. See "Photometry and Illumination," § (93) *et sqq.*; also "Spectrophotometry," *passim.*

High Potential Generators for X-rays. See "Radiology," § (15).

Hilger's Constant-deviation Spectrometer, used as monochromatic illuminator. See "Immersion Refractometry," § (7).

Hilger's Sector Photometer: an instrument used in the photographic method of spectrophotometry. See "Spectrophotometry," § (17).

Högner's Method of calculating Average Illumination. See "Photometry and Illumination," § (70).

Holders, Plate-, for portable field and copying cameras. See "Photographic Apparatus," § (5) (i.).

Hollow Glass-ware, Manufacture of. See "Glass," § (18) (i.).

Homo-hetero-analysis (of colour): the matching of a colour by a mixture of grey with the correct proportion of homogeneous radiation of the correct wave-length. See "Spectrophotometry," § (3).

Horn, E♭ Tenor: a brass wind-instrument with valves. See "Sound," § (44).

French, in C; scale on, tabulated. See *ibid.* § (38), Table VIII.

French, with valves. See *ibid.* § (38).

French, without valves. See *ibid.* § (37).

Hot-cathode Tube: a tube in which electrons are produced from an electrically heated spiral of tungsten. The vacuum in the tube is so high that the residual gas plays no active part. See "Radiology," § (8).

Hue: that attribute of colour in light by virtue of which it differs from grey. See "Spectrophotometry," § (2); see also "Eye," § (8).

Hüfner's Spectrophotometer. See "Spectrophotometry," § (12).

Humours: the fluid contents of the eye. See "Eye," § (2).

Huygens' Construction: a graphical method of determining the form of an advancing wave-front by considering it as the envelope of the secondary wavelets emitted by all points on the original wave-front. See "Polarised Light and its Applications," § (6).

Huygens' Eyepiece. See "Telescope," § (6); "Eyepieces," § (2).

— I —

IMMERSION REFRACTOMETRY

§ (1) INTRODUCTION. — The ordinary direct methods[1] of measuring the refractive indices of glass specimens cannot be applied when one wishes to determine the refractive properties of a specimen of glass which is unpolished, or is in the form of a lens, or is very small. It is frequently either inadvisable or impossible to grind and polish optically flat surfaces on a specimen, on account of its being a component of some optical system which it is essential to keep intact, or because of its small size. In such cases the usual method adopted for measuring the refractive index for any given wave-length is to immerse the specimen in a transparent liquid having approximately the same index, and to vary the concentration of the liquid until the refractive indices of liquid and specimen for that wave-length are the same. The refractive index of the liquid is then measured by means of one of the ordinary methods.

Four of the more important methods which depend on this principle are as follows :

§ (2) CHRISTIANSEN'S METHOD.—Christiansen[2] described an immersion method of measuring the refractive index of glass in the form of a fine powder. The powder is placed in a hollow glass prism of refracting angle 45° to 60°, which is filled with a mixture of carbon disulphide and benzene. When a homogeneous mixture has been obtained by means of stirring, the prism is placed on the table of a spectrometer, the collimator slit of which is illuminated by a sodium flame. The sodium line is seen sharply defined in the telescope. The powder gradually sinks to the bottom of the prismatic cell and the constitution of the mixture becomes different in different horizontal layers. As a result of this a band of light is seen in the field of view of the telescope. After some minutes, however, the powder collects at the bottom of the cell. One now sees two sharp lines, one due to refraction in the clear liquid and the other due to refraction in the saturated mixture. The refractive index of the powder may then be determined from the results of measurements of the refractive indices of liquid and mixture for two slightly different concentrations of the liquid. Christiansen determined by this method the refractive indices of a specimen of crown glass for the lines C, D, and F. His mean results differed from those obtained by direct measurement by 1 to 5 units in the fourth decimal place.

The above method can only be applied to the case of very finely divided powders, for with coarser powders homogeneous mixtures cannot be satisfactorily obtained. Christiansen measured the refractive properties of coarser powders by a different method, which is easier to apply, but does not yield quite so accurate results. The principle of the method is as follows. If one introduces glass powder into a suitable mixture of carbon disulphide and benzene, one can observe two colours, one of which consists of light which passes through without refraction, while the other is composed of all the other colours. The two colours are therefore complementary, the first being monochromatic and the second heterochromatic. Christiansen in his experiments made use of two sensitive tints corresponding to monochromatic light of wave-lengths 530 and 460 $\mu\mu$; the heterochromatic colours were reddish-violet and dirty yellow respectively. The concentration of the carbon disulphide and benzene mixture is gradually altered until the first sensitive tint is observed. The refractive indices of the liquid and the

[1] See article on "Spectroscopes and Refractometers."

[2] C. Christiansen, *Wied. Ann.*, 1884, xxiii. 298; 1885, xxiv. 439.

glass are then the same for the wave-length 530 $\mu\mu$. The indices of the liquid for the lines C, D, and F are determined in the ordinary way, and the index for the wave-length 530 $\mu\mu$ is obtained by interpolation or with the aid of Cauchy's dispersion formula. The process is then repeated for the second sensitive tint. In this way the refractive indices of the glass for the wave-lengths 530 and 460 $\mu\mu$ are determined. A specimen of flint glass was experimented with, and the values of the indices obtained by this method for the critical wave-lengths differed from those derived by direct measurement by 8 units and 1 unit in the fourth decimal place respectively.

§ (3) Chalmers' Method.—Chalmers [1] employed a method of measuring the refractive indices of a lens, which consisted in immersing the lens in a liquid of approximately the same index and determining its focal length. If μ_1 is the index of the lens and μ_0 that of the liquid, the focal length F is given by

$$F=(\mu_1-\mu_0)(R_1-R_2),$$

where R_1, R_2 are the curvatures of the surfaces, considered positive when convex to the incident light, provided that the thickness of the lens is small in comparison with the radii of curvature. It is then only necessary to determine the values of F, R_1, and R_2 to the degree of accuracy that is required for the value of $\mu_1-\mu_0$. A series of stable transparent liquids of suitable refractive index is used; for example, cedar oil ($\mu_D=1{\cdot}517$ approximately) and oil of cloves ($\mu_D=1{\cdot}530$ approximately) are useful for the measurement of the indices of crown lenses. An average accuracy of about 5 units in the fourth decimal place was obtained with the method.

§ (4) Martin's Method. — The principle of this method [2] is to immerse the specimen in a suitable liquid, such as carbon disulphide or mercury potassium iodide, contained in a prismatic cell resting on the prism table of a spectrometer, and to adjust the concentration of the liquid until the spectrum line, for which the refractive index of the specimen is to be measured, is seen in sharp focus. The angle of minimum deviation for the line in question is then determined and the refractive index calculated in the usual way. The diluent employed is alcohol, and the liquid in the cell is kept homogeneous by mechanical stirring. An average accuracy of about 1·4 in the fourth decimal place was obtained by this method. Measurements were made on the C, D, and F lines of the spectrum, but the author states that when using mercury potassium iodide as the immersion liquid it was not easy to determine the values for the F line.

§ (5) Cheshire's Method. — An accurate method of immersion refractometry has been described by R. W. Cheshire.[3] It is based on the shadow method introduced by Foucault [4] for figuring surfaces. The image of a vertical straight edge, backed by a source of monochromatic light, is projected by a telescope objective into the plane of a second straight edge, which is so arranged as to cover about half the full aperture of the objective of an observing telescope. A rectangular cell containing the glass specimen immersed in a suitable liquid (mercury potassium iodide was employed) is placed near the first objective on the side away from the source of light. The strength of the liquid is varied until, on traversing the second straight edge across the objective of the observing telescope, which is focussed on the glass specimen, the whole field darkens simultaneously and uniformly. This only happens when the refractive indices of the glass and the liquid are the same. The index of the liquid is measured directly, the cell being cemented to the block of a Pulfrich refractometer. An average of 2 units in the fifth decimal place was obtained by this method in measurements on the sodium D line. It was found difficult to obtain accurate values of the indices for the hydrogen C and F lines owing to want of sufficient brightness.

One great disadvantage in most of the methods so far described is the necessity for accurately adjusting the strength of the liquid so as to obtain equality of index for any given standard wave-length. On approaching the balance point it is frequently found that, after adding a drop of the concentrated liquid or of the diluting medium, in order, as the case may be, to increase or decrease the density of the mixture, the point of equality has been overreached. Even when one goes to the trouble of preparing for each wave-length two mixtures, one slightly denser and the other slightly less dense than the mixture of required refractive index, a considerable time is spent before the correct amounts of these mixtures have been added. Then again, even if the correct density of the immersion liquid has been attained, it may alter appreciably, due to evaporation at the surface, during the time occupied in making the necessary measurements of the refractive index. Such difficulties are overcome in the two following methods, descriptions of which have recently been published.

[1] S. D. Chalmers, *Proc. Opt. Convention*, 1905, p. 198.

[2] L. C. Martin, *Opt. Soc. Trans.*, 1916, xvii. 76.

[3] R. W. Cheshire, *Phil. Mag.*, 1916, xxxii. 409.

[4] L. Foucault, *Ann. de l'Observatoire de Paris*, 1859, v. 197; *Recueil des travaux scientif.*, Paris, 1878, 232; see article on "Objectives, The Testing of Compound."

§ (6) Fabry's Method.—This method[1] depends on the measurement of small differences of refractive index by comparison with the index of a reference prism. The immersion liquid is contained in a cell with approximately parallel sides, and the reference prism and the specimen whose index is required are simultaneously placed in the liquid. A certain number of standard reference prisms is required (Fabry uses five), and in each case that one is chosen whose index is nearest to that of the specimen under examination. The index of the liquid is then determined by means of the deviation which the prism produces relatively to the ray which traverses the cell without passing through the prism. Fabry employs carbon disulphide as the immersion liquid. The cell is placed between a collimator and a telescope mounted on a divided circle, which reads to about one second. The faces of the cell are set normal to the collimator beam, and the reference prism, whose refracting angle is 90°, is placed in the cell with its faces equally inclined to the incident beam. The observing telescope is provided with an objective of 40 cm. focal length and an adjustable eyepiece with millimetre graduations. The collimator slit is illuminated by means of a source of light giving a line spectrum. Now if A be the angle and n the refractive index (corresponding to the wave-length of one of the lines) of the reference prism, and N the index of the liquid, the angle of deviation D due to the prism is given by

$$D = 2(n - N)\tan A/2, \quad . \quad . \quad (1)$$

if the difference between n and N is small.

Thus, if A=90°,

$$N = n - D/2, \quad . \quad . \quad . \quad (2)$$

an equation which gives N in terms of the known refractive index n and the deviation D, which is measured on the circle. If D is greater than 1° it is advisable to use a more accurate formula, such as

$$N = n - \tfrac{1}{2}\sin D + \frac{1}{8n}\sin^2 D + \frac{1}{4n^2}\sin^3 D + \frac{15}{128n^3}\sin^4 D + \ldots, \quad (3)$$

of which the third term is very nearly quite negligible. It gives correct values, where A=90°, up to values of D of about 15°.

The method of obtaining the refractive indices of a specimen in the form of a lens is as follows: The strength of the liquid is roughly adjusted so that its index is very nearly equal to that of the lens for a given radiation. The position of the telescope draw-tube, for which the line in question is sharply focussed, is determined when the reference prism alone is immersed in the liquid. The lens is then immersed in the path of the beam of light, and the position of the draw-tube is altered by an amount x so as to bring the line into sharp focus again. At the same time the deviation D is measured. The composition of the liquid is then slightly altered and new values of x and D are measured. Having obtained two or three pairs of values, including positive and negative values of x, a curve is drawn with x and D as co-ordinates. From this curve, which is practically a straight line, the value of D which corresponds to $x=0$—that is, to equality of index of lens and liquid—can be deduced. The index of the lens can then be obtained from the formula (3). Fabry finds that the error of focussing the telescope is of the order of ± 1 mm. and theoretically deduces the fact that, in order to obtain for refractive indices an accuracy of one in the fifth decimal place, the difference in thickness of the centre and of the edge of the effective portion of the lens must at least reach the value 6000λ. Thus a difference in thickness of 4 mm. is sufficient to ensure this accuracy throughout the spectrum.

The method as applied to a specimen in the form of a prism is somewhat similar. The prism is immersed in such a way that its edge is perpendicular to the edge of the reference prism. A wire is stretched across the collimator slit, its image appearing as a black point. The vertical deviation d of this image is measured by means of a micrometer eyepiece, the motion of which is parallel to the length of the slit, and at the same time the horizontal deviation D caused by the reference prism is determined. The simultaneous measurements of D and d are made for two or three different strengths of the immersion liquid, and a curve is drawn through the points corresponding to the values obtained. From this curve, which is practically a straight line, the value of D which corresponds to $d=0$ may be obtained, and then the index of the specimen can be deduced, as in the former case, from formula (3). Fabry found that the method gave results correct to 1 or 2 units in the fifth decimal place in the case of a specimen of quartz, the angle between adjacent faces of which was 120°. He does not indicate for what wave-length the measurement was made.

The method can also be applied to the case of irregular fragments with curved faces (such as rods of glass or beads of borax fused to platinum wire) and transparent substances in the form of powders with coarse grains.

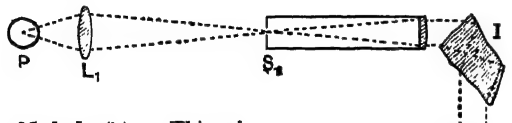
The method is in many respects a great improvement on those that have already been described, but it is doubtful whether the degree of accuracy claimed, namely one unit in the fifth decimal place of refractive index, will be attained in general practice.

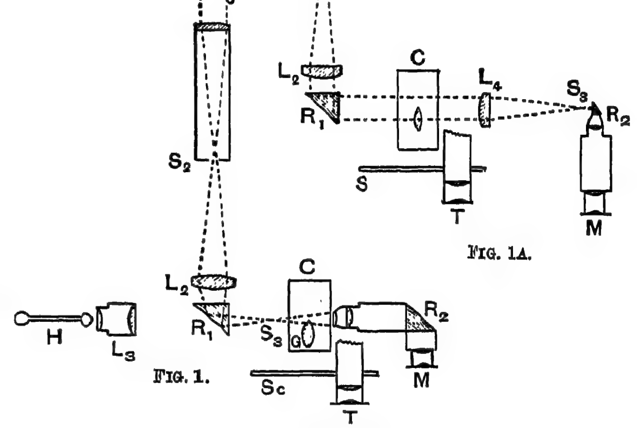
[1] C. Fabry, *Journ. de Physique*, 1919, xi. 11.

§ (7) ANDERSON'S METHOD.—In this case[1] the use of auxiliary prisms of known refractive index is not required. The salient point of the method consists in the substitution of a variation of wave-length for a variation of the strength of the immersion liquid. The apparatus required consists of a rectangular glass cell for holding the immersion liquid,[2] a Hilger constant deviation spectrometer used as a monochromatic illuminator, a Pulfrich refractometer for measuring the index of the liquid, and suitable observing apparatus for determining the wave-lengths for which the indices of the liquid and the specimen are equal. A number of alternative methods are used for finding the balance points; they are as follows:

Method (i.).—This is illustrated in plan in *Fig.* 1. The light from a "Pointolite" tungsten arc P is focussed on the first slit S_1 of the monochromatic illuminator I by means of a lens L_1. An image of a vertical spider-line stretched across the second slit S_2 of the illuminator is formed at S_3 by the lens L_2, a reflecting prism R_1 being used to deflect the beam of light through a right angle. C represents the cell containing the liquid in which the specimen of glass G is immersed. The mechanical stirrer used for mixing up the liquid in the cell is not shown. Optical contact is made between the base plate of the cell and the block of the Pulfrich refractometer by means of a drop of α-monobromonaphthalene, or, in the case of extra dense flints, mercury barium iodide. T represents the observing telescope, Sc the vertical scale, and L_3 the condensing lens of the refractometer. H is the hydrogen vacuum tube which is used as the source of light for the measurement of the refractive index of the liquid. The telescope T during the experiment is below the horizontal plane containing the axis of the optical system for measuring the balance point of refractive index. A low-power microscope M is focussed on S_3 through the liquid. A 3-inch objective is used in the microscope in order to get sufficient working distance between S_3 and the microscope for the introduction of the cell. For convenience of observation a reflecting prism R_2 is employed to deflect the beam of light in the microscope through a right angle. In this way the eye-pieces of the refractometer telescope and the microscope are brought close together. The prism R_1 is mounted on a small table attached to the refractometer and is rotated out of the position shown in the diagram when the wave-length has been adjusted for equality of index of liquid and glass, in order that the refractometer measurement may be made.

The method of measuring the refractive index of the glass specimen G for, say, the hydrogen C line is as follows: The concentration of the liquid is first of all roughly adjusted until its refractive index for C is nearly equal to that of the specimen. This can be rapidly done, before the cell is placed in position on the refractometer block, by altering the strength of the liquid until the image of an object (preferably illuminated by red light) seen through the specimen and the liquid very nearly coincides with the image seen through the liquid alone. When the cell is in position, the specimen is mounted in such a way that it intercepts part of the beam of light which passes through the liquid. In general, therefore, two images of the spider-line will be seen in the microscope; the one which is formed by the light which traverses the specimen and the liquid will be fairly well defined only if the specimen is regular in shape (such as a lens or prism). The drum of the monochromatic illuminator is now rotated until the two images are brought into coin-





FIGS. 1 and 1A.

[1] J. S. Anderson, *Opt. Soc. Trans.*, 1920, xxi. 105.
[2] Mercury potassium iodide is used.

cidence and the scale reading *s* is noted. The prism R_1 is then swung out of position and the scale reading *r* on the micrometer drum of the Pulfrich refractometer, corresponding to the refractive index of the liquid *for the C line,* is determined. The reading *s* can be made from a position in the neighbourhood of T and M by placing a reflecting prism and lens combination above the drum of the monochromatic illuminator, the scale being illuminated by a flash-lamp bulb.

The concentration of the liquid is next altered slightly by introducing a drop of liquid of somewhat higher or lower refractive index, as the case may be, and the observer waits until the mixture is made homogeneous by stirring. Two images of the spider-line will again be seen, and their separation will depend on the effective angle of the edge of the specimen used and on the difference of index of the liquid caused by the addition of the drop. In order to get as great a separation as possible it is therefore advisable to adopt such an orientation of the specimen as to present a maximum effective angle in the path of the beam. The wave-length of the light is now altered until the scale readings *s* and *r* are determined, the latter corresponding to the refractive index of the new mixture *for the C line.* The process is repeated for two or three different concentrations of the liquid, its refractive index being varied by small steps.

The pairs of values of *s* and *r* thus obtained are now plotted and a curve is drawn through the points. From this curve the value of *r*, which corresponds to the scale reading *s* for the wave-length of the hydrogen C line, is determined; this value of *r* then gives the angular reading on the refractometer which would have been obtained if a mixture with the same refractive index as the specimen for the C line had been made up. The reason for plotting values of *s* and *r*, and not the values of the wave-lengths and the refractive indices, is that the curve thus obtained *is a straight line,* provided that the scale of the monochromatic illuminator is an evenly divided one. On this account only a few sets of readings are necessary for each determination of refractive index.

A similar process to the above is applied in the case of the other wave-lengths employed, namely those corresponding to the sodium D_1 and hydrogen F lines. In the case of the former the refractometer settings are made for the sodium D_2 line, and a correction has to be made in order to obtain the index for the D_1 line.

Method (ii.).—The second method that was adopted for determining the balance points is similar to the one just described, but is somewhat more sensitive. The optical system employed is shown in *Fig.* 1A. The image S_3 of the vertical spider-line stretched across the second slit S_2 of the monochromatic illuminator is formed by two lenses L_2 and L_4, one on either side of the cell C, the beam of light which passes through the liquid being rendered parallel by the lens L_2. One advantage of using this system rather than the one shown in *Fig.* 1 is that a higher power microscope may be employed, since its working distance is not limited by the interposition of the cell between S_3 and the microscope objective. A small right-angle prism R is attached to the objective of the microscope M so as to bring the eyepiece of the latter into close proximity to that of the refractometer telescope T.

Both of these methods (but especially the latter) are very convenient for observing the balance point, particularly when the specimen under examination is in the form of a lens or prism. Attention may again be directed to the fact that, in order to obtain the greatest accuracy with any given specimen, the greatest effective angle of the specimen should be utilised, so as to give a maximum deflection of the image formed by the light which passes through the glass and the liquid. Thus, for example, in the case of a flat lens the specimen should be mounted edgewise in the beam of light, for with such an orientation it is practically equivalent to a thin prism of large angle.

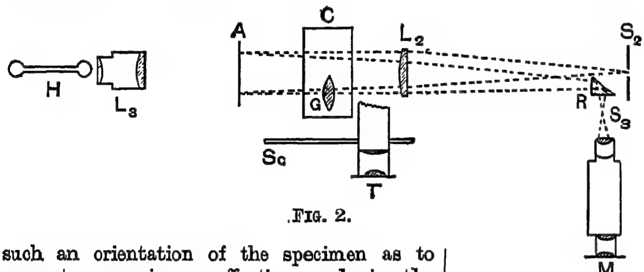


FIG. 2.

A modification of method (ii.), whereby the beam of light is made to pass twice through the specimen, and thus give double the deviation, is shown in *Fig.* 2. The second slit S_2 of the monochromatic illuminator is placed at the focus of the lens L_2, so that a parallel beam of light is caused to pass through the cell C. This beam is reflected back by a plane mirror A, slightly inclined to the beam, and after again traversing the cell is brought to a focus at S_3, a small right-angle prism R being inserted so as to deflect the light through 90°. The spider-line image formed at S_3 is

then observed by means of a microscope M. The mirror A may be mounted on a table which can be swung out of position when the refractometer measurements are being made, or it may be fixed in such a way that the beam of light from the hydrogen tube H passes underneath it. By placing the mirror at a small angle to the beam, the reflected light is brought to a focus at the side of the slit S_2, thus involving less loss of light than if a half-silvered mirror were used instead of the prism R, and, further, the images formed by light reflected from the walls of the cell do not disturb the main image which is to be observed.

Method (iii.).—Another method, which is convenient to employ in the case of lenses and prisms, the edges of which form fairly large effective angles, is illustrated in *Fig.* 3. A intersect corresponds then to the wave-length for which the refractive indices of the liquid and the specimen are equal. This method is a very pleasant one to work with, but it does not give such accurate results as the previous ones, unless the angle of the specimen is fairly large. With specimens of small angle it is sometimes rather difficult to determine the point C with accuracy, as the region round C is filled with systems of fringes.

Method (iv.).—Another method of observing the balance point is similar to the one employed by R. W. Cheshire, and is especially convenient in the case where the specimen to be examined is in the form of an irregular fragment. The general arrangement of apparatus is the same as that shown in *Fig.* 1A, except that, instead of using a fine vertical spider-line across the second slit of the monochromatic illuminator,

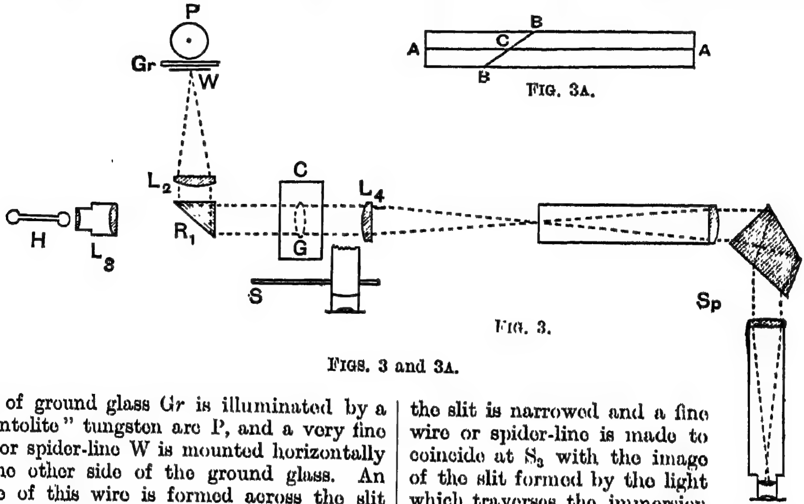


FIG. 3A.

FIG. 3.

FIGS. 3 and 3A.

piece of ground glass *Gr* is illuminated by a "Pointolite" tungsten arc P, and a very fine wire or spider-line W is mounted horizontally on the other side of the ground glass. An image of this wire is formed across the slit of a spectrometer *Sp* by means of two lenses L_2 and L_4, the wire being at the focus of the former. The cell C is mounted in the path of the parallel beam of light between L_2 and L_4, the general arrangement of apparatus being much the same as in the previous method. The specimen is mounted in such a way that its edge intercepts the upper half of the beam of light which passes through the liquid in the cell. Before the specimen is immersed in the liquid one sees in the spectrometer a spectrum traversed horizontally by a fine dark line AA (*Fig.* 3A), corresponding to the point on the slit where the image of the wire W falls. If, now, the specimen is placed in position, another dark line BB will be observed. This line, which is straight if the specimen is a prism and approximately straight if it is a lens, makes an angle with the line AA depending on the effective angle of the specimen. The point C where these two lines the slit is narrowed and a fine wire or spider-line is made to coincide at S_3 with the image of the slit formed by the light which traverses the immersion liquid alone. No observing microscope is required; on placing one's eye behind the wire at S_3 no light reaches it unless the refractive indices of the liquid and the specimen are different. The method of observation is, therefore, to alter the wave-length of the light until the specimen suddenly disappears in the field of view. This modification of the shadow method does not, of course, give such accurate results as Cheshire's method, but it has the advantage of greater compactness and, in combination with the rest of the process, of greater available intensity.

Method (v.).—One of the most accurate methods of observing the balance point is to make use of an interferometer system.[1] For the sake of simplicity this may be built up

[1] The problem has recently been dealt with by C. Barus, *The Interferometry of Reversed and Non-reversed Spectra*, Part II., Carnegie Institution of Washington Publication, 1917, ccxlix. 95.

on the plan of the Rayleigh interference refractometer.[1] Preliminary experiments carried out some years ago with a Michelson interferometer system showed that very accurate determinations of the equality of refractive index could be obtained. The method, however, necessitates the employment of an optically accurate, parallel-sided cell, and the system must be mounted in such a way that it is free from vibration. For most purposes, therefore, the increased sensitivity does not compensate for the more rigorous conditions required. The method would, however, be extremely useful for detecting and measuring small variations of refractive index in specimens of irregular shape.

In experiments with prisms having angles of about 12·5° the results obtained by using method (ii.) were found to give for the C, D_1, and F lines a mean error of about 5 units in the fifth decimal place, though this error may be somewhat reduced by modifications of the optical systems employed.

In conclusion, it is interesting to note that the principle of immersion refractometry is used in mineralogy for determining the refractive indices of mineral fragments.[2] J. S. A.

Indirect Lighting. See "Photometry and Illumination," § (71).

Indoor Lighting. See "Photometry and Illumination," § (71).

Induction Coil: an open-core, high-tension static transformer with independent interrupter. See "Radiology," § (16).

Infra-red, wave-length measurements in the. See "Wave-lengths, The Measurement of," § (7).

INFRA-RED TRANSMISSION AND REFRACTION DATA ON STANDARD LENS AND PRISM MATERIAL[3]

§ (1) Introduction.[4]—This paper gives exact data on the spectral transparency and, in particular, the refractivity of materials which are useful for prisms and lenses for spectroradiometers.

The data on refractive indices were taken from smooth curves drawn through values which were collected from various sources and reduced to a common temperature.

It is important, especially in work of the highest precision (such as, for example, the determination of the constant of spectral radiation), to use the most precise instruments and optical data available. It is therefore relevant to discuss very briefly some recent designs of optical instruments suitable for spectroradiometry.

The pioneering investigation of the infra-red refractive indices of a substance dates back to 1886 when Langley determined the dispersion of rock salt to about 5μ. In these determinations a spectrometer having an image-forming mirror of long focal length was used. In subsequent determinations of the refractive indices of rock salt and of fluorite, the image-forming mirror of the spectrometer used by Langley had a focal length of 4 to 4·7 metres. The apparatus was in a large enclosure which could be maintained at a constant temperature. Langley (19) was therefore justified in calling attention to the very high precision attainable, "owing, if to no other reason, to the far greater size of the apparatus employed, where size is a most important element of accuracy." Other experimenters (20, 23, 24), using his methods but having spectrometer mirrors of only about one-twelfth the focal length, have attempted to produce similar data which, unfortunately, have been given the widest recognition in tables of physical constants. These published results, especially the older ones, have been very confusing to the writer who, for some years, has been confronted with the task of obtaining reliable refractive indices.

The recent measurements on rock salt (26) and on fluorite (28) by Paschen, when corrected for temperature (33, 34), are in agreement with Langley's (19) measurements. The numerical values, given in the present paper, have been adopted after a careful study of all the data available.

§ (2) The Spectroradiometer. — For measuring thermal radiation intensities in the ultra-violet part of the spectrum one may use a spectrometer having achromatic lenses of quartz-fluorite. However, the scarcity of clear fluorite for large-sized lenses makes such apparatus very expensive.

Pflüger (1) used simple lenses of fluorite 4 cm. in diameter (32 cm. focal length) and a fluorite prism. An inexpensive spectroradiometer of high light-gathering power was made by Coblentz (2) by using simple plano-convex lenses (6 cm. in diameter and 20 cm. focal length) and a prism of quartz. Pfund (5) has described similar apparatus in which the radiometer is kept in focus automatically in different parts of the spectrum.

The apparatus may be designed also as an illuminator for separating the visible from the ultra-violet of, for example, the sun or a quartz mercury vapour lamp (4).

For spectroradiometric measurements in the visible spectrum and the infra-red to about

[1] Lord Rayleigh, *Phil. Mag.*, 1917, xxxiii. 161.

[2] H. A. Miers, *Mineralogy*, 1902, 259, 260; F. E. Wright, *Journ. Washington Acad. of Sc.*, 1914, iv. 389; E. S. Larsen, *U.S. Geol. Survey*, Bulletin 679, 1921.

[3] With special reference to Infra-red Spectroradiometry.

[4] References are given in a classified bibliography at the end of this paper.

$0{\cdot}8\mu$ we can use a spectrometer with visually achromatised lenses of glass. Here, also, it is desirable to use apparatus having a high light-gathering power, such as one obtains with lenses 6 cm. in diameter and 20 cm. focal length (2, 3, 4).

A common property of all metals is a low reflectivity in the ultra-violet and in the violet-blue part of the visible spectrum (15, 16, 17). Furthermore, the spectral reflectivity in the short wave-length is greatly reduced on tarnishing of the metal. Hence concave mirrors of metals have never been used extensively for spectroradiometric measurements in the visible and in the ultra-violet spectrum.

Because of the lack of achromatism (and the opacity of the material) lenses of glass, quartz, fluorite, etc., achromatised for the visible spectrum, have not been used extensively in infra-red spectroradiometric work. Lehmann (46) has described an infra-red spectrograph achromatised for $\lambda = 0{\cdot}589\mu$ and $\lambda = 1{\cdot}529\mu$.

A concave mirror is achromatic (also astigmatic) and hence spectrometers with collimating and image-forming mirrors, instead of lenses, have been used almost exclusively for infra-red spectral radiation intensity measurements. In the infra-red spectrum beyond 2μ most of the metals have a very high reflecting power (90 to 98 per cent), and concave metal mirrors, or metal-on-glass mirrors are, therefore, especially useful for infra-red investigations.

Recent designs of spectrometers having collimating and image-forming mirrors are described in papers by Coblentz (6) and by Gorton (9). Vacuum spectrometers have been described and used by Trowbridge (7) and by McCauley (8).

In order to obtain the spectral energy distribution of an incandescent substance it is necessary to correct the observations for absorption of radiation by the mirrors and by the prism. The proper formula for eliminating the absorption in a wedge is given by Paschen (29), and by Coblentz (39), who gives also the numerical factors for eliminating the absorption in a wedge of quartz.

It is beyond the scope of this paper to discuss the construction and operation of the instruments (bolometers, thermopiles, etc.) used for measuring the thermal radiation intensities. References are given in the appended Bibliography (10) on "Radiometers."

§ (3) Spectrometer Calibration.—In most spectroradiometric work it is necessary to know the wave-lengths at which the thermal radiation intensities are measured. In the visible spectrum it is an easy matter to note the spectrometer settings for the emission lines of some source (*e.g.* the mercury arc or helium gas in a Plücker tube), the wave-lengths of whose emission lines are known. Similarly in the ultra-violet the emission lines of mercury, cadmium, zinc, etc., may be noted with a fluorescent canary glass screen, or radiometrically with a thermopile (1) or bolometer.

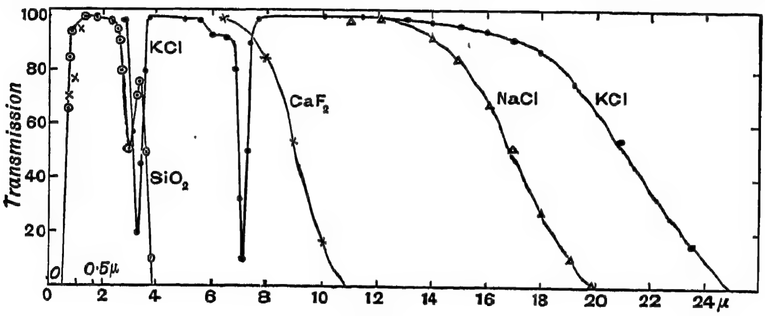


Fig. 1.

The spectrometer circle may be calibrated for wave-lengths in the infra-red spectrum to 1μ by noting the emission lines (12) of sodium and potassium in a carbon arc; also the emission lines of a quartz mercury vapour lamp and helium in a vacuum tube.

Beyond 2μ, where the emission lines are usually weak (except the strong emission band of carbon dioxide at $4{\cdot}4\mu$ in the bunsen flame) one can calibrate the prism by noting sharp absorption bands (13, 14) such as, for example, the bands of sylvite, KCl, illustrated in *Fig.* 1.

For work requiring great accuracy the proper method of calibration is by calculating the minimum deviation settings for different wave-lengths, using the refractive indices and the angle of the prism. For this purpose the yellow sodium lines, or, better, the yellow helium line, $\lambda = 0{\cdot}5875\mu$, is used as a reference point on the spectrometer circle. The minimum deviation settings for the various infra-red wave-lengths are computed from the corresponding refractive indices, and referred

to the yellow helium line as a basis. After this the bolometer or thermopile is adjusted upon the yellow helium line, then on rotating the spectrometer through a certain angle, say 2°, the corresponding wave-length is, say, 6μ, while a rotation of 4° places the bolometer at about $8{\cdot}7\mu$ in the spectrum of a 60° fluorite prism ((6), p. 49).

Elimination of Scattered Radiation in Spectral Energy Measurements.—In the design of optical instruments there are opportunities for great improvement in this respect. Take, for example, the image-forming telescope of a spectrometer. The telescope tube should be large and suitably diaphragmed so that when the violet end of the spectrum is incident upon the radiometer receiver the infra-red end of the spectrum cannot be reflected from the side of the tube and impinge upon the receiver. Furthermore, the bevelled edges of the exit slits of the spectrometer should face outwards (18) instead of facing the image-forming lens, as obtains in commercial instruments.

By using suitably constructed optical instruments the scattered radiation is practically eliminated. What little remains may be obviated by using, before the entrance slit of the spectrometer, a shutter (22, 5, 18), which is opaque to the region of the spectrum under investigation, but which transmits the scattered radiations. In this manner the scattered radiations are incident upon the radiometer all the time and, hence, do not affect the energy measurements. Using a spectrometer which is provided with slits and diaphragms, as just mentioned, it has been found (4) that the scattered radiation was immeasurable in comparison with the intensities under investigation.

§ (4) Optical Constants of Glass.—In the ultra-violet end of the spectrum ordinary crown glass is transparent to about $0{\cdot}3\mu$, while the flint-silicate glasses absorb strongly throughout the blue and violet end of the spectrum.

In the infra-red spectrum all glasses (38, 40) begin to absorb at about 2μ, and for a thickness of 1 cm. they are practically opaque to radiations of wave-length greater than 3μ. Glasses containing traces of iron impurities have an absorption band at 1μ.

The refractive indices of various glasses have been determined by Rubens (20). He determined the refractive indices also of water, xylol, benzol, etc. However, in view of the fact that the refractive indices depend upon the composition of the glass, no refraction data are given in this paper. Practically no infra-red work is being done with glass prisms.

§ (5) Optical Constants of Carbon Disulphide.—Carbon disulphide is quite transparent in the infra-red. In the region to 3μ Rubens (20) found an absorption of only 5 to 7 per cent for a 1 cm. thickness. Beyond 4μ there are a number of very large absorption bands (36).

As illustrated in *Fig.* 3, carbon disulphide has a very much larger dispersion than quartz, etc., in the region of 0·5 to 2μ, and hence is especially adapted for certain fields of spectro-radiometry.

The infra-red refractive indices of carbon disulphide were determined by Rubens (20). Those in the visible and in the ultra-violet were determined by Flatow (32). Rubens's values of the infra-red refractive indices are given in Table I.

Table I

Indices of Refraction of Carbon Disulphide in Air at 15° C. (Rubens)

Wave-lengths. μ=0·001 mm.	Refractive Index, n.	Log n.
$0{\cdot}434\mu$	1·6784	·2248955
0·485	1·6550	·2187980
0·590	1·6307	·2123741
0·656	1·6217	·2099705
0·777	1·6104	·2069338
0·823	1·6077	·2062050
0·873	1·6049	·2054480
0·931	1·6025	·2047980
0·999	1·6000	·2041200
1·073	1·5978	·2035224
1·164	1·5960	·2030329
1·270	1·5940	·2024883
1·396	1·5923	·2020249
1·552	1·5905	·2015337
1·745	1·5888	·2010692
$1{\cdot}998\mu$	1·5872	·2006317

§ (6) Optical Constants of Quartz.—Quartz is one of the most useful materials for prisms. It is extremely transparent to ultra-violet radiations. Pflüger (41) found a transmission of 94 per cent at $0{\cdot}222\mu$ and 67 per cent at $0{\cdot}186\mu$, for a sample of crystalline quartz 1 cm. in thickness. Some samples of amorphous quartz have been found to be more opaque than crystalline quartz; but this may be the result of contamination in melting.

The infra-red transmission of quartz has been determined by various observers. A characteristic absorption band occurs at about $2{\cdot}95\mu$. A sample 1 cm. in thickness is practically opaque (38) to radiations of wave-length greater than 4μ (see *Fig.* 1).

The absorption, reflection, and dispersion constants of quartz are given in a paper by Coblentz (39), who determined the transmission of samples 3 cm. in thickness. The paper gives also factors for eliminating the absorption in a quartz prism.

In the short wave-lengths the refractive indices of quartz have been determined by

Martens (31). In the infra-red there are important determinations by Rubens (21, 23), Carvallo (35), and Paschen (27). Carvallo's data extend to 2·2μ, and in the region of 1·45 to 1·8μ they are slightly lower (by several units in the 5th decimal place) than three determinations made by Paschen. In this region of the spectrum the data must therefore be considered uncertain with the doubt in favour of Carvallo's data. This uncertainty affects Warburg's (43) determination of the spectral radiation constant by perhaps 0·2 to 0·4 per cent.

The refractive indices (ordinary ray) of quartz at 18° C. are given in Table II. They are taken from a graph of sufficient size to permit reading the data to 1 or 2 units in the fifth decimal place. In many cases the values agree exactly with Carvallo's measurements. Paschen's data may be recognised by the fact that his wave-lengths are given to the fifth decimal place.

§ (7) Optical Constants of Fluorite.—Fluorite is very transparent to radiations of wave-lengths extending from 0·2μ to 10μ (see *Fig.* 1). Pflüger (41) found a transmission of 86 per cent at 0 23μ and 70 per cent at 0·186μ, for a sample 1 cm. in thickness. Lyman (42) examined fluorites from various sources, and of various colours, and found that they are opaque to radiation of wave-lengths less than about 0·12μ. Coblentz (38) examined green fluorites with a view of determining their suitability for prisms. He found numerous sharp absorption bands, in the infra-red, which would render such material unsuitable for prisms.

The refractive indices of fluorite have been determined by various observers (21, 23), and repeatedly by Paschen (23, 24, 25, 28). Applying temperature coefficients of refraction (33, 34), it is found that Paschen's determinations, especially the latest ones (28), which were obtained with an improved spectrometer, coincide with the dispersion curve of fluorite determined, to 3·5μ, by Langley (19). Beyond 4μ the dispersion of fluorite is much larger than at 1·5 to 2μ, and there is better agreement among the various determinations of the refractive indices. Furthermore, slight deviations have less effect upon spectral radiation measurements.

In Table III. the refractive indices of fluorite to 3·5μ are taken from the smooth curve published by Langley (19). In many cases Paschen's original wave-lengths are retained. As already mentioned, the corresponding refractive indices, when corrected for temperature of the prism, fall exactly upon Langley's curve of refractive indices.

Beyond 3·5μ the refractive indices are taken from the smooth curve (extended from 3μ) through the various determinations of Paschen

Table II

Indices of Refraction of Quartz in Air at 18° C. (Carvallo, Paschen)

Wave-lengths. μ = 0·001 mm.	Refractive Index, n.	Log n.
0·54609μ	1·54617	·1892573
0·58758	1·54430	·1887317
0·58932	1·54424	·1887148
0·61577	1·54323	·1884306
0·66784	1·54154	·1879548
0·6731	1·54139	·1879126
0·6950	1·54078	·1877407
0·70654	1·54048	·1876561
0·72817	1·53995	·1875066
0·7711	1·53895	·1872245
0·8007	1·53834	·1870523
0·8325	1·53773	·1868801
0·84467	1·53752	·1868208
0·8671	1·53712	·1867078
0·9047	1·53649	·1865298
0·9460	1·53583	·1863432
0·9914	1·53514	·1861480
1·01406	1·53486	·1860405
1·0417	1·53442	·1859443
1·08304	1·53390	·1857970
1·0973	1·53366	·1857291
1·12882	1·53328	·1856214
1·1592	1·53283	·1854940
1·17864	1·53263	·1854373
1·2288	1·53192	·1852361
1·3070	1·53090	·1849468
1·3195	1·53076	·1849071
1·3685	1·53011	·1847226
1·3958	1·52977	·1846162
1·4219	1·52942	·1845268
1·47330	1·52879	·1843478
1·4792	1·52865	·1843081
1·4972	1·52843	·1842427
1·52961	1·52800	·1841234
1·5414	1·52782	·1840722
1·6087	1·52687	·1837021
1·6146	1·52680	·1837822
1·6815	1·52585	·1835118
1·7487	1·52486	·1832300
1·76796	1·52462	·1831616
1·8487	1·52335	·1827997
1·9457	1·52184	·1823690
2·0531	1·52005	·1818579
2·06262	1·51991	·1818178
2·1719	1·51799	·1812689
2·35728	1·51449	·1802664
2·3840	1·51400	·1801259
2·4810	1·51200	·1795518
2·575	1·51100	·1792645
2·65194	1·50824	·1784704
2·79927	1·50474	·1774614
3·09393μ	1·49703	·1752305

(25) and Rubens (20). All these data are reduced to 20° C.

Langley's data are referred to the "A line," $\lambda = 0·7604\mu$. The most convenient reference point for adjusting a spectroradiometer in the spectrum is the yellow helium line, $\lambda = 0·58758\mu$. Until recently, when Paschen (28) determined

TABLE III

INDICES OF REFRACTION OF FLUORITE IN AIR AT 20° C. (LANGLEY, PASCHEN, RUBENS)

Wave-length.	Refractive Index, n.	Log n.
0·48615μ H_β	1·43704	·1574695
0·58758 He	1·43388	·1565120
0·58932 Na	1·43384	·1564995
0·65630 H_α	1·43249	·1560916
0·68671	1·43200	·1569430
0·72818 He	1·43143	·1567601
0·76653 K	1·43093	·1566184
0·88400	1·42980	·1552753
1·0140 Hg	1·42884	·1549835
1·08304 He	1·42843	·1548589
1·1000	1·42834	·1548316
1·1786	1·42789	·1546948
1·250	1·42752	·1545822
1·3756	1·42689	·1543905
1·4733	1·42642	·1542474
1·5715	1·42596	·1541073
1·650	1·42558	·1539916
1·7680	1·42502	·1538210
1·8400	1·42468	·1537173
1·8688 He	1·42454	·1536747
1·900	1·42439	·1536274
1·9153	1·42431	·1536046
1·9644	1·42407	·1535313
2·0582 He	1·42360	·1533880
2·0626	1·42357	·1533789
2·1608	1·42306	·1532232
2·250	1·42258	·1530766
2·3573	1·42198	·1528936
2·450	1·42143	·1527255
2·5537	1·42080	·1525329
2·6519	1·42018	·1523460
2·700	1·41988	·1522517
2·750	1·41956	·1521538
2·800	1·41923	·1520528
2·850	1·41890	·1519518
2·9466	1·41823	·1517467
3·0500	1·41750	·1515231
3·0980	1·41714	·1514128
3·2413	1·41610	·1510939
3·4000	1·41487	·1507134
3·5359	1·41376	·1503788
3·8306	1·41119	·1495855
4·000	1·40963	·1491051
4·1252	1·40837	·1487476
4·2500	1·40722	·1483620
4·4000	1·40568	·1478864
4·6000	1·40357	·1472341
4·7146	1·40233	·1468502
4·8000	1·40130	·1465311
5·000	1·39908	·1458426
5·3036	1·39522	·1446427
5·8932	1·38712	·1421141
6·4825	1·37824	·1393312
7·0718	1·36805	·1361020
7·6612	1·35675	·1324998
8·2505	1·34440	·1285290
8·8398	1·33075	·1240965
9·4291μ	1·31605	·1192724

the refractive index of this line, there has been some uncertainty in infra-red spectral radiation measurements (44) requiring the highest accuracy. An error of 5″ (or $n = 3 \times 10^{-5}$) in the determination of this refractive index affects the constant of spectral radiation by 0·3 per cent.

The dispersion curve of fluorite is illustrated in *Figs.* 2 and 3 (from Rubens).

In the infra-red the variation of the refractive

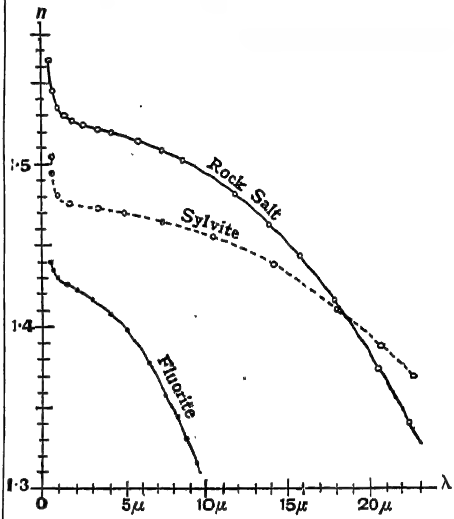


FIG. 2.

index of fluorite (34) with temperature decreases slowly with wave-lengths. At 1μ the coefficient of variation amounts to about $\Delta n = 0{\cdot}000012$ and at 6·5μ it amounts to about $\Delta n = 0{\cdot}000009$ for 1° rise in temperature.

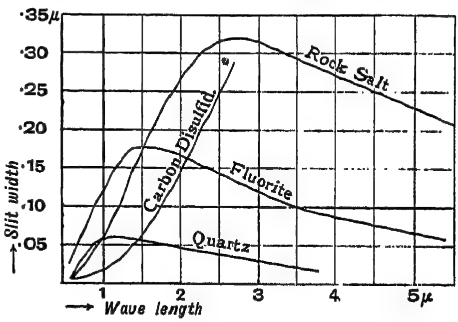


FIG. 3.

§ (8) OPTICAL CONSTANTS OF ROCK SALT.—Rock salt is uniformly transparent from 0·2μ in the extreme ultra-violet (41) to 12μ in the infra-red (23) (see *Fig.* 1). In the region of 15μ the absorption increases rapidly. A plate of rock salt 1 cm. in thickness is completely opaque (22) to radiation of wave-lengths greater than 20μ. The refractive indices of

rock salt have been determined in the short wave-lengths by Martens (30, 31), and in the infra-red by Langley (19), by Rubens (21, 22), and by Paschen (26). In the region of 1 to 3μ there is considerable disagreement among the older determinations. However, the recent work of Paschen (26) is in excellent agreement with Langley's measurements, which are, without doubt, very accurately determined.

The infra-red refractive indices of rock salt, at 20°, are given in Table IV. The first part of the table, to 5μ, consists principally of Langley's (and Paschen's corrected for temperature) measurements as read from the smooth curve ((19), p. 235, Plate XXIX.). Beyond 5μ, to 16μ, the refractive indices are principally Paschen's measurements corrected for temperature (34); also some of Rubens's measurements and several interpolated values.

The temperature coefficient of refraction of rock salt (19, 33, 34) decreases slowly with wave-length; amounting to about $\Delta n = 0{\cdot}000038$ at 1μ and $\Delta n = 0{\cdot}000025$ at 9μ for 1° rise in temperature.

The general outline of the dispersion curve of rock salt is illustrated in *Fig.* 2 (from Rubens).

§ (9) Optical Constants of Sylvite.—Of all the substances which are otherwise suitable for prisms, sylvite, KCl, is transparent throughout the greatest part of the infra-red spectrum. A plate 1 cm. in thickness transmits (22) radiations to 24μ (see *Fig.* 1). In the region of 5μ to 10μ the dispersion is small. Furthermore, there are sharp absorption (11, 13) bands at $3{\cdot}18\mu$ and $7{\cdot}08\mu$. Hence sylvite is the most useful for investigations in the region of the spectrum extending from 10 to 20μ.

In the short wave-lengths the refractive indices of sylvite have been determined by Martens (31). In the infra-red we have various determinations by Rubens (22) (with Nichols and with Trowbridge), by Trowbridge (45), and by Paschen (26).

The infra-red refractive indices of sylvite are given in Table V. They are read from a smooth curve (practically Paschen's curve) drawn through the various determinations, all of which are in close agreement, except at 9 to 11μ, where the older determinations do not agree very well with Paschen's data.

The temperature coefficient of refraction of sylvite observed by Liebreich (34) decreases from

$$\Delta n = 0{\cdot}0000364 \text{ at } \lambda = 0{\cdot}589\mu \text{ to}$$

$$\Delta n = 0{\cdot}000031 \text{ at } \lambda = 8{\cdot}85\mu.$$

§ (10) Summary: Comparison of Dispersive Materials.—In *Fig.* 3 is given the width that a radiometer receiver, of 4′ of arc, subtends in wave-lengths, in different parts of

Table IV

Indices of Refraction of Rock Salt in Air at 20° C. (Langley, Paschen, Rubens)

Wave-length. μ=0·001 mm.	Refractive Index, n.	Log n.
0·5893μ	1·54427	·1887232
0·6400	1·54141	·1879182
0·6874	1·53930	·1873233
0·7604	1·53682	·1866230
0·7858	1·53607	·1864110
0·8835	1·53395	·1858112
0·9033	1·53361	·1857149
0·9724	1·53253	·1854090
1·0084	1·53206	·1852758
1·0540	1·53153	·1851255
1·0810	1·53123	·1850404
1·1058	1·53098	·1849695
1·1420	1·53063	·1848702
1·1786	1·53031	·1847794
1·2016	1·53014	·1847312
1·2604	1·52971	·1846091
1·3126	1·52937	·1845126
1·4874	1·52845	·1842512
1·5552	1·52815	·1841660
1·6368	1·52781	·1840693
1·6848	1·52764	·1840238
1·7670	1·52736	·1839414
2·0736	1·52649	·1836960
2·1824	1·52621	·1836142
2·2464	1·52606	·1835716
2·3560	1·52579	·1834947
2·6505	1·52512	·1833040
2·9466	1·52466	·1831730
3·2736	1·52371	·1829024
3·5359	1·52312	·1827341
3·6288	1·52286	·1826600
3·8192	1·52238	·1825231
4·1230	1·52156	·1822891
4·7120	1·51979	·1817836
5·0092	1·51883	·1815092
5·3009	1·51790	·1812432
5·8932	1·51593	·1806792
6·4825	1·51347	·1799738
6·80	1·51300	·1795518
7·0718	1·51093	·1792443
7·22	1·51020	·1790345
7·59	1·50855	·1785597
7·6611	1·50822	·1784647
7·9558	1·50665	·1780124
8·04	1·6064	·1779403
8·8398	1·50192	·1766468
9·00	1·50100	·1763807
9·50	1·49980	·1760333
10·0184	1·49462	·1745308
11·7864	1·48171	·1707632
12·50	1·47568	·1689922
12·9650	1·47160	·1677898
13·50	1·4666	·1663117
14·1436	1·46044	·1644837
14·7330	1·45427	·1626450
15·3223	1·44743	·1605976
15·9116	1·44090	·1586338
17·93	1·4149	·1507257
20·57	1·3735	·1378287
22·3μ	1·3403	·1272020

the spectrum produced by prisms of carbon disulphide, quartz fluorite, and rock salt. These data are required for reducing the spectral energy distribution from the prismatic into the normal spectrum (6).

TABLE V

INDICES OF REFRACTION OF SYLVITE IN AIR AT 15° C. (PASCHEN, TROWBRIDGE, RUBENS)

Wave-length. μ=0·001 mm.	Refractive Index, n.	Log n.
0·5893μ	1·49044	·1733145
0·656	1·48721	·1723723
0·7858	1·48328	·1712232
0·845	1·48230	·1709361
0·884	1·48142	1706782
0·9822	1·48008	·1702862
1·003	1·47985	·1702177
1·1786	1·47831	·1697655
1·584	1·4765	·1692335
1·7680	1·47595	·1690717
2·3573	1·47475	·1687184
2·9466	1·47388	·1684621
3·5359	1·47305	·1682164
4·125	1·47215	·1679521
4·7146	1·47112	·1676481
5·3039	1·47001	·1673202
5·50	1·46962	·1672050
5·8932	1·46880	·1669627
6·50	1·46750	·1665781
7·00	1·46625	·1662080
7 661	1·46450	·1656894
8·00	1·46350	·1653927
8·2505	1·46272	·1651642
8·8398	1·46086	·1646086
9·500	1·45857	·1639273
10·0184	1·45672	·1633760
10·500	1·45475	·1627883
11·00	1·45263	·1621550
11·786	1·44919	·1611253
12·50	1·44570	·1600782
12·965	1·44346	·1594048
14·144	1·43722	·1575232
15·00	1·4320	·1559430
15·912	1·42617	·1541713
16·50	1·42230	·1529912
17·00	1·41885	·1519365
17·680	1·41403	·1504586
18·10	1·4108	1494655
19·00	1·4031	·1470886
20·00	1·3939	·1442316
20·60	1·3882	·1424520
22·5μ	1·3692	·1364669

From these curves it is evident that, in the region of 0·5 to 1·5μ, a carbon disulphide prism is the most useful for producing a large dispersion.

The next best prism material is quartz, which is the most useful in the region of the spectrum extending from the visible to 2·8μ in the infra-red. Beyond this point a quartz prism is too opaque for practical work.

From the standpoint of dispersion and transparency, a fluorite prism is the most useful in the region of 2μ to 9μ. However, the material is difficult to obtain, and the next best substance is rock salt, which permits measurements to 14μ when using a 60° prism, and to 16μ when using a 30° prism. By enclosing the spectrometer (14) and by keeping the prism covered when not in use, the faces of a rock-salt prism are easily protected from moisture.

There are few but sylvite prisms in existence, and their usefulness is confined to that part of the spectrum extending from 10 to 20μ.

BIBLIOGRAPHY

Spectroradiometers

1. Pflüger, *Ann. der Phys.*, 1904, iv. 13, p. 890.
2. Coblentz, *Bur. Stds. Bulletin*, 1911, vii. 245.
3. Coblentz, *Bur. Stds. Bulletin*, 1913, x. 38 (Fig. 5).
4. Coblentz and Kahler, *Bur. Stds. Bulletin*, 1920, xvi. 233.
5. Pfund, *Phys. Rev.*, 1916, ii. 7, p. 289.
6. Coblentz, *Bur. Stds. Bulletin*, 1913, x. 1.
7. Trowbridge, *Phil. Mag.*, 1910, vi. 20, p. 768.
8. McCauley, *Astrophys. Jour.*, 1913, xxxvii. 164.
9. Gorton, *Phys. Rev.*, 1916, ii. 7, p. 66.

Radiometers

10. For bibliography and description of bolometers, thermopiles, etc., see papers by Coblentz, *B.S. Bull.*, 1907, iv. 391; 1912, ix. 7; 1914, xi. 131; 1916, xiii. 423; and 1918, xiv. 507.

Emission and Absorption Bands

11. Rubens and Trowbridge, *Ann. der Phys.*, 1895, iii. 60, p. 724.
12. Paschen, *Ann. der Phys.*, 1908, iv. 27, p. 537; 1909, 29, p. 625, and 30, p. 746; 1910, 33, p. 717.
13. Coblentz, *B.S. Bulletin*, 1911, vii. 652.
14. Coblentz, Publ. Nos. 35, 65, and 97, Carnegie Inst., Washington.

Reflecting Power of Metals

15. Hagen and Rubens, *Ann. der Phys.*, 1901, iv. 1, p. 352; 1902, viii. 1.
16. Coblentz, *B.S. Bulletin*, 1910, vii. 197; 1917, xiv. 115.
17. Hulburt, *Astrophys. Jour.*, 1915, xlii. 205.

Elimination of Scattered Radiation

18. Coblentz, *Jour. Franklin Inst.*, 1913, clxxv. 151; *B.S. Bulletin*, 1918, xiv. 229 and 534. See also Rubens (11) and Pfund (5).

Absorption and Refraction of Lens and Prism Material

19. Langley, *Annals Astrophys. Obs. of the Smithsonian Inst.*, i. 222, plate xxvi. B, 1902.
20. Rubens, *Ann. der Phys.*, 1892, iii. 45, p. 238.
21. Rubens, *Ann. der Phys.*, 1895, iii. 51, p. 381; 54, p. 476.
22. Rubens (with Nichols and with Trowbridge), *Ann. der Phys.*, 1897, iii. 60, pp. 418 and 724; 61, p. 225.
23. Paschen, *Ann. der Phys.*, 1894, iii. 53, pp. 301, 337, and 812.
24. Paschen, *Ann. der Phys.*, 1895, iii. 56, p. 762.
25. Paschen, *Ann. der Phys.*, 1901, iv. 4, p. 299.
26. Paschen, *Ann. der Phys.*, 1908, iv. 26, pp. 120 and 1029.
27. Paschen, *Ann. der Phys.*, 1911, iv. 35, p. 1005.
28. Paschen, *Ann. der Phys.*, 1913, iv. 41, p. 670.
29. Paschen, *Sitzber. Akad. Wiss.*, Berlin, 1899, xxii. 405.
30. Martens, *Ann. der Phys.*, 1901, iv. 6, p. 603.
31. Martens, *Ann. der Phys.*, 1902, iv. 8, p. 459.
32. Flatow, *Ann. der Phys.*, 1903, iv. 12, p. 85.
33. Micheli, *Ann. der Phys.*, 1902, iv. 7, p. 772.

34. Liebreich, *Verh. Phys. Gesell.*, 1911, xiii. 1.
35. Carvallo, *Compt. Rend.*, 1898, cxxvi. 728.
36. Coblentz, Publ. No. 35, Carnegie Inst. of Washington, p. 16, Fig. 50, 1905.
37. Coblentz, *B.S. Bulletin*, 1911, vii. 652.
38. Coblentz, *B.S. Bulletin*, 1912, ix. 112.
39. Coblentz, *B.S. Bulletin*, 1914, xi. 471.
40. Coblentz, Emerson and Long, *B.S. Bulletin*, 1918, xiv. 653.
41. Pflüger, *Phys. Zs.*, 1904, v. 215.
42. Lyman, *Astrophys. Jour.*, 1906, xxiii. 197.
43. Warburg and Müller, *Ann. der Phys.*, iv. 48, p. 410; *Zeitsch. für Instrk.*, 1915, xxxv. 98.
44. Coblentz, *B.S. Scientific Papers*, 1920, No. 406, p. 7.
45. Trowbridge, *Ann. der Phys.*, 1898, iii. 65, p. 595.
46. Lehmann, *Ann. der Phys.*, 1912, iv. 39, p. 62.

W. W. C.

INSTRUMENTS, CHIEF MUSICAL, COMPASSES OF. See "Sound," § (25).

INTEGRATING HEMISPHERE PHOTOMETER. See "Photometry and Illumination," § (120).

INTEGRATING PHOTOMETERS: devices for obtaining the average candle-power of a light source by means of a single measurement. See "Photometry and Illumination," § (45) *et sqq.*

INTENSITY OF A SPECTRUM, ESTIMATION OF THE. See "Spectroscopy, Modern," § (4).

INTENSITY OF X-RAYS, chemical methods of measuring. See "Radiology," § (25).

Fluorescence methods of measuring. See *ibid.* § (25).

Ionisation methods of measuring. See *ibid.* § (25).

Methods of measuring. See *ibid.* § (25).

Photographic methods of measuring. See *ibid.* § (25).

INTERCARDINAL ROLLING ERROR. See "Navigation and Navigational Instruments," § (13) (ii.).

INTERFEROMETER, LUMMER AND GEHRCKE. See "Light, Interference of," § (11).

INTERFEROMETER, MICHELSON'S. See "Light, Interference of," § (9).

Twyman's modification for optical testing. See "Interferometers, Technical Applications," § (5).

INTERFEROMETER METHODS FOR THE DETERMINATION OF STANDARD WAVE-LENGTHS. See "Wave-lengths, The Measurement of," § (3).

INTERFEROMETERS: TECHNICAL APPLICATIONS

I. INTERFERENCE FRINGES

§ (1) INTRODUCTORY.—In utilising interference methods in Optical Test work we are always concerned with nearly parallel plates, or with arrangements which are optically equivalent to them. The character of the interference fringes obtainable from such plates depends on the method of illumination and the method of observation. They can conveniently be divided into three classes.

CLASS I.—If a plate is illuminated by a broad source of light[1] and is viewed by the naked eye from a convenient distance, the rays which reach the eye from different points of the plate are reflected from the surfaces of the latter at different angles. The path difference between the rays reflected from the two surfaces at any point $=2\mu t \cos r$, where r is the angle of incidence on the *back* surface, μ the refractive index, and t the thickness of the plate. For a complete treatment of the fringes produced under such circumstances the reader should consult Mann's *Manual of Advanced Optics* or Michelson (*Phil. Mag.*, 1898). For our present purpose it is sufficient to note that the fringes will be the loci of points for which $2\mu t \cos r$ is constant, and therefore indicate variations of μ, t, or r jointly or separately. If t is very small, *i.e.* for extremely thin films, the variation of phase due to the variation of r at different points of the film may be negligible, so that if the optical thickness μt is uniform the film will appear uniformly dark or bright all over. If either μ or t vary, fringes will be seen which are the loci of equal optical thickness.

If the thickness is appreciable, however, the variation of phase due to the varying angle of incidence becomes of importance, and for any but quite thin plates the fringes are practically loci of points from which the light reaching the eye meets the plate at equal angles of incidence. With a plane parallel plate they are circles concentric with the normal from the eye to the plate and are located at infinity. If the surfaces are not parallel, the fringes are still circles or arcs of circles, but their centres are displaced from the normal. If the surfaces are irregular, or if the material of the film is not perfectly homogeneous, so that from one or both of these causes there are local variations of μt, there will be local distortion of the circular fringes at these points. Such irregularities will only be noticeable, however, near the centre of the system where the fringes are not too close together. With any but the thinnest films the fringes become so closely packed, as the incidence angle increases, that they are only visible in the neighbourhood of normal incidence.

To summarise, therefore, fringes of Class I. are formed in such circumstances that the light corresponding to each point in the fringe system reaches the eye from separate points (or relatively small regions) of the film, and is reflected from the latter at varying

[1] Except where otherwise specified, the light is assumed to be monochromatic.

angles. Except for extremely thin films, the fringes are mainly loci of equal incidence angles, being very slightly affected in shape and position by variations in film thickness.

CLASS II.—Instead of viewing the fringe system at infinity with the naked eye, as supposed in the preceding discussion, we may employ a telescope of large aperture focussed for infinity, so that for all parts of the field of view light from the whole surface of the film is received by the telescope. The fringes are then purely loci of equal inclination, and are unaffected in shape or position by variations in the optical thickness of the film. Thus if the telescope is directed normally to the film, circular fringes concentric with the axis will be seen in the focal plane of the eyepiece. Since, at any point in this plane, light from the whole of the film is focussed, the phase at any such point depends on the *mean* film thickness and the angle of incidence of the beam. Any variations in film thickness contribute their effect equally to all points in the fringe system, which therefore indicates the inclination of the rays only, and tells us nothing about the flatness, homogeneity, or parallelism of the film.

If, however, μt varies much, the *distinctness* of the fringes will be impaired; for we can regard the resultant illumination at the focal plane of the telescope in such a case as due to the superposition of a number of exactly similar fringe systems, of which some differ in phase from others. The result of this is to diminish the contrast between the bright and dark parts of the field.

Thus, to summarise, fringes of Class II. are formed when light corresponding to each point of the fringe system comes from the whole of the film (or from the same part of it), but meets it at varying angles of incidence. They are purely fringes of equal inclination. If they are crisp and distinct we deduce that the film is of fairly uniform optical thickness over the region utilised; but no clue whatever is given to the character or position of any variations which may actually be present.

CLASS III.—If we employ a point source at infinity, such, for example, as a small illuminated pinhole at the focus of a collimating lens, all rays strike the film at the same angle of incidence. There are consequently no variations of phase from one part of the film to another except such as may be due to variations of μt. To observe a fringe system under such circumstances it is necessary to employ a telescope with the eyepiece removed. The object-glass collects the parallel beam after it leaves the film and produces an image of the pinhole in its focal plane. If the eye is placed at this image, the whole surface of the film (if the object-glass is large enough) is seen illuminated and traversed by fringes which are true contours of the optical thickness, μt.

These fringes are not located at a definite distance from the eye, as are those of Classes I. and II. They are visible at all distances, and so appear to coincide with any surface on which the eye is focussed.[1]

If the film is thick the fringes will be most distinct at perpendicular incidence. It is not practicable to use a theoretical point source: considerations of brightness require that the hole shall have an appreciable area, and its image in the focal plane of the telescope lens may be regarded as a small portion of the Class II. fringe system produced by a broad source. If this portion is in the centre, where the phase varies slowly with the angle of incidence, there will be no phase difference between the light from one part of the pinhole and another; but if it is at an outer part of the system where the Class II. fringes are closely packed, the small cone of rays which reach the eye from any point of the film comprises an appreciable range of phase retardation, and the contour fringes are rendered indistinct.

II. TECHNICAL APPLICATIONS

§ (2) TEST PLATES.[2]—The simplest application of interference fringes is the testing of optical surfaces by means of test plates. The surface under test is placed in contact with another surface which it ought to fit perfectly. This may either be plane, for prism surfaces, etc., or may be curved, for lens surfaces. A separate test plate is of course necessary for every curvature. The surface and test plate are worked together to squeeze out the air between them and obtain a very thin film. This is viewed by the light from a window or bright wall. Fringes of Class I. are obtained, hence the necessity to work the film down as thin as possible in order that the fringes should truly represent film contour rather than variation of incidence angle. With white light the fringes are coloured owing to the variation of retardation with wave-length. If the surfaces are very close and quite parallel, the film will appear of a uniform colour, which will vary with the obliquity from which it is examined. Variations of thickness insufficient to produce a complete fringe will cause a variation of tint; and if the general tint is yellow or bluish green, the two regions of the spectrum where tint varies most rapidly with wave-length, the test is very sensitive. It is usually possible to secure a sensitive tint by working the film and choosing a suitable obliquity.

§ (3) CONSTANCY OF OPTICAL THICKNESS OF

[1] Guild, "Location of Interference Fringes," *Phys. Soc. Proc.*, 1920, xxxii. 32.

[2] See also "Optical Parts, Working of," § (4).

Flat Discs.—This is a most important test. For many purposes, for instance in making echelon gratings, Lummer Gehrke plates, etc., it is essential to have plates of which the variations of optical thickness from one part to another are reduced to a very small amount indeed.

The best known arrangement for testing such plates is shown in *Fig.* 1 (*a*). Light from a broad source S is thrown normally on the plate AB by a glass plate MM′. The fringes are observed by an eye at E. The fringes will be of Class I. and will be concentric circles located at infinity if the glass is nearly parallel. As we have seen, the form of these fringes does not at once tell us much about the variations of optical thickness. To determine this, the plate is moved parallel to itself. If the optical thickness is uniform, the fringes will not move; but if it varies, a fringe will pass across any point, E′ say, on the fixed support, for every half wave-length by which the optical thickness varies. By moving the plate in such a way that no fringes pass E′, it is possible to trace out contour lines of equal optical path.

If desired, a *small* telescope with cross-lines in the eye-piece may be used at E.

This method is tedious and troublesome, since each contour line has to be traced separately by indirect means. By a simple modification of the arrangement the plate can be examined by fringes of Class III., and the fringe system itself is then the contour map. A lens LL′, *Fig.* 1 (*b*), is interposed between MM′ and AB, and a point source (an illuminated pinhole) is placed at the focus of LL′. A parallel beam of rays meets the plate AB. It is reflected back through LL′ and refocussed in the neighbourhood of S. Some light is reflected at the surfaces of MM′ to form two images of the source at *e* and *e*′. MM′ should be thick so as to separate *e* and *e*′ as widely as possible. An eye placed at *e* or *e*′ sees the plate AB traversed by interference fringes which are the required contours of equal optical thickness.

For the reasons given earlier, the fringes will only be distinct with thick plates provided the incidence is normal. The table on which AB rests should be provided with levelling screws. The normal adjustment is then easily made, because the retardation is a maximum in this position, and on altering the tilt the fringes will move towards a certain configuration and then retreat from it as the perpendicular position is passed.

Care must of course be taken to ensure that the light is properly collimated. This is done by adjusting so that the reflected image,

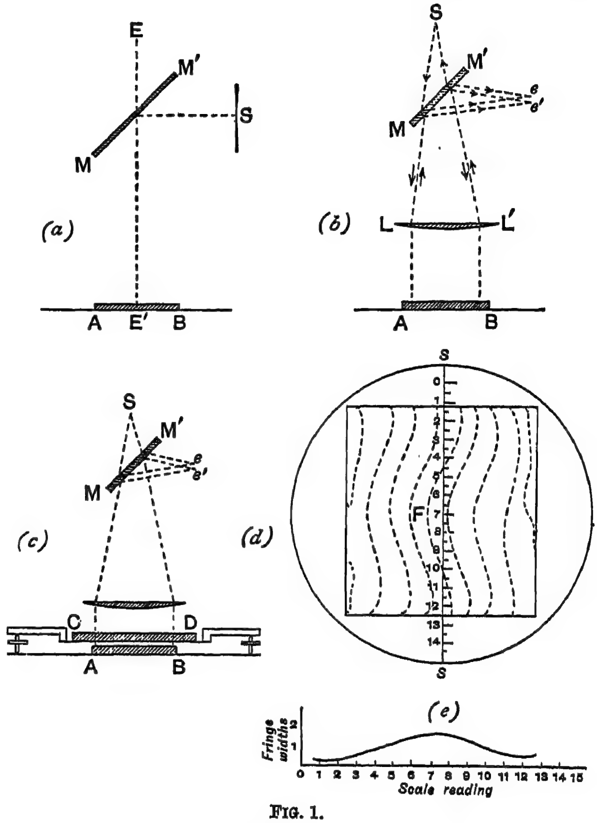


Fig. 1.

when formed just beside S, is in sharp focus.

When a suitable source of monochromatic light is used [1] plates of over an inch thick may be examined in this way.

§ (4) Contour Maps of Flat Surfaces.—The same apparatus affords a very convenient means of determining the contour maps of surfaces which are intended to be flat. The customary method of working the surface into close contact with a test plate and examining the fringes with a broad source of light is unsatisfactory on various grounds. In the first place, troubles sometimes arise, due

[1] Guild, *Phys. Soc. Proc.*, 1920, xxxii. 341.

to the surfaces not being perfectly dry, in which case local patches of perfect blackness, due to absence of air film, may be formed. Further, it is possible, when two nearly flat surfaces are wrung together in this way, to squeeze out slight convexities and make the surface under test appear better than it really is. Lastly, in spite of the greatest care to exclude dust particles, it frequently happens that the surfaces get scratched in the process.

It is on every ground more satisfactory to test the surfaces without putting them in contact at all. If the upper surface of the slab AB, *Fig.* 1 (*c*), is the surface to be tested, the test plate is mounted above it, with its standard face downward, supported on a suitable carrier with three adjustment screws. It is clear that in general there will be four reflected beams, each of which will give rise to images of S in the neighbourhood of *e* and *e′*. If any two of the four surfaces are parallel the corresponding images will coincide. It is desirable that the test plate CD should be somewhat prismatic, so that the image from its upper surface may be out of the way. By means of the levelling screws the test plate is adjusted so that those images which are respectively due to the upper surface of AB and the lower surface of CD coincide. If the eye is then placed at *e* or *e′* fringes are seen which are the loci of equal thickness of the air film. By carefully adjusting the test plate while watching the fringes, the latter can be broadened out until, if the film is perfectly parallel, a uniform illumination is obtained. If the surfaces are not perfectly flat there will be a position in which a minimum number of fringes appear, which form a contour map of the joint defects of the surfaces at all points. This is the best adjustment in order to see at a glance the general character of the contour; but to make quantitative measurements it is better to proceed indirectly and use a slightly wedge-shaped film such that the fringes are about a centimetre or so apart. A diametral scale, *ss*, divided to ½ centimetres, should be drawn in ink on the surface of the test plate, and also a line on the surface under test passing approximately through the centre, which should be marked by means of a dot. The lower block is then moved on the table until this line coincides with the line on the test plate as seen by the eye at *e*. The position of the central dot on the scale is noted. The fringes are then adjusted until they run as nearly parallel to the scale as possible and are about a centimetre or so wide. The appearance of the field of view might then be something like *Fig.* 1 (*d*). The most convenient fringe, say F, may be taken as the datum of film thickness, and the thickness at different distances along the scale *ss* may be specified by the distance of the scale from this datum fringe, as measured in fringe widths. For instance, with a configuration as drawn (if the thickness of the film is increasing from left to right) the excess of thickness above the datum is ·4 at 3 on the scale, 1·05 at 4·5, 1·6 at 6, 1·7 at 7, 1 at 10, ·6 at 12, and so on. If these are plotted to a suitable scale, a contour line of the variations of film thickness along the line *ss* is obtained (*Fig.* 1 (*e*)), the unit corresponding to one interference fringe, that is, to a difference in thickness of one ½ wavelength.

The surface under test is then rotated until another line, making, say, 30° with the first, coincides with the scale on the test plate, and the contour along this line is determined in the same way. This is done for a number of radial lines and also for two non-radial lines intersecting the others as shown in *Fig.* 2 (*a*).

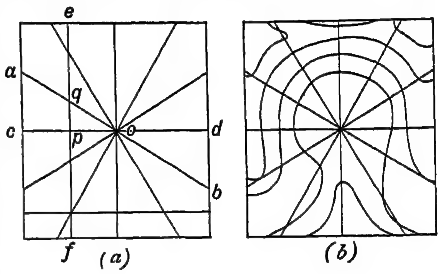


FIG. 2.

The non-radial contours give relations between second points on each of the radial contours, which enable us to alter the position and inclination of the base lines of the latter so as to make the contours mutually consistent. Thus we may take the contour along *ab* as first obtained: the base line for any other contour, such as *cd*, must be altered so that the ordinates corresponding to the common point O are equal, and also so that the ordinate at *p* differs from the ordinate of the *ab* contour at *q* by the amount indicated by the non-radial contour along *ef*. If now an outline of the surface is drawn on paper, and along each radial line we mark off the distances from O at which the ordinates of the revised contours have particular values (differing, say, by $\frac{1}{10}$th fringe), we obtain a contour map of the surface, such as *Fig.* 2 (*b*). It will be observed that in the whole process we are only concerned with the properties of the test plate along one line *ss* (*Fig.* 1 (*d*)), and the contour of the plate along this line is easily determined. If two other fairly good surfaces are available, by testing one line of each of these against the test plate and against each other, their contours and that of the test plate can be deduced. The method is thus absolute and no surface has to be taken for granted.

§ (5) TESTING BY TRANSMITTED LIGHT.—*F. Twyman* [1] has evolved an elegant method of testing prisms, lenses, and even complete optical instruments by interference. The instrument employed is the well-known Michelson interferometer.[2] The general principles of this instrument are described in all text-books. It is only necessary to remind readers that the instrument is optically equivalent to two approximately parallel planes separated by an air film of a thickness which can be varied from zero to any desired value by moving one of the mirrors. This being so, we can obtain with the instrument fringes of any of the three classes previously described. The most suitable type to employ depends on the purpose in view.

If a broad source is employed, and the fringes are viewed by the naked eye, fringes of Class I. will be obtained. If a telescope is employed the fringes will be of Class II. These are the arrangements most commonly described in text-books and employed in the physical laboratory; but they are quite unsuitable for optical testing. For this purpose fringes of Class III. are required, and the arrangement of the interferometer to give these is shown in *Fig.* 3 (*a*). Light from a suitable source S is focussed by means of a condensing lens on a small "pinhole" P which is at the focus of a lens L_1. We have thus the equivalent of a very small source at infinity. That portion of the parallel beam which is reflected by the half-silvered surface of the oblique plate m strikes the mirror M_1 normally, and returns along its own path until it again meets the oblique plate. Here a portion is reflected back through L_1 and refocussed at P, and a portion transmitted. The latter is focussed in an image of P at P′ by a lens L_2. Similarly a portion of the beam, which is in the first instance transmitted by m, forms, after reflection by M_2 and partial reflection at m, a second image of P in the neighbourhood of P′. If the image of M_2 by reflection in m is very nearly parallel to M_1 the two images at P′ will coincide. An eye placed at P′ will then see interference fringes of Class III., which are the contour lines of the thickness of the "air film" to which the instrument is equivalent. If M_2 is carefully adjusted while examining the fringes, they can be broadened out until the air film is parallel, when, if the surfaces are perfectly flat, the film will appear of uniform brightness. If the surfaces are not perfectly flat (or, rather, if their irregularities are not exactly similar and similarly situated) the film can never be cleared of fringes over its whole area, and the fringe system, when broadened out as much as possible, forms a contour map of the irregularities.

It is clear, of course, that whether the fringes have been broadened out or not, they give precisely the same information regarding the irregularities of the film; but they are

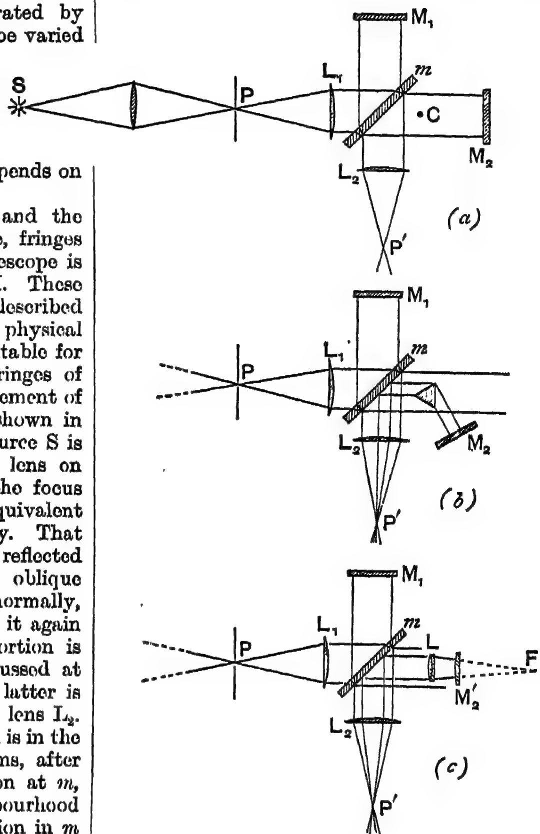


FIG. 3.

equivalent to the contour map of a country based on a datum plane inclined to the horizontal. If the inclination were considerable such a map would consist of closely packed, nearly straight lines parallel to the horizon, hills and valleys appearing simply as slight bends in the contours. Such a map would give all the information as to the nature of the country that could be deduced from an ordinary one; but it is much less easy to interpret its general appearance. Similarly, whatever the adjustment of the mirrors, the fringe system gives full information as to their departures from exact similarity of surface; but it is usually convenient to

[1] F. Twyman, *Phil. Mag.*, Jan. 1918, p. 49; *Photographic Journal*, Nov. 1918, p. 230; and *Astrophys. Journ.*, 1918, xlviii. 256.

[2] See also "Wave-lengths, Measurement of," § (3).

have the fringes as broad as possible in order to render them easily interpretable. Sometimes, however, an adjustment departing slightly from parallelism is helpful in deciding between two possible interpretations.

In the interferometer, as used for optical testing, the mirror M_1 is mounted on a carriage which rests in a geometrical slide. By pulling one or other of two strings the mirror can be moved nearer to or farther from the oblique plate without its inclination to the direction of the light being seriously altered. The mirror M_2 is arranged on a carriage which can rotate about a point C. Each mirror is provided with delicate adjustment screws for final adjustment of the parallelism of the equivalent film. The mirrors are worked to such accuracy that when used alone, as in *Fig.* 3 (*a*), a field free from fringes can be obtained.

The method of testing an optical piece, such as a slab or a prism, *or any optical instrument of which the focal length is infinite*, consists in placing it in the beam mM_2, and adjusting M_2 to be perpendicular to the transmitted rays so that they are returned through the instrument and finally focussed at P′. The eye at P′ then sees the aperture of the instrument traversed by a fringe system which is the contour map of the wave front, originally plane, after passing twice through the instrument. The general method will be fully understood from its application to a spectroscope prism. The prism is placed in the position indicated in *Fig.* 3 (*b*). That portion of the beam which is transmitted by the prism is deviated in the direction shown. M_2 is adjusted so as to reflect this beam back through the prism, after which it is partially reflected by m and focussed at P′. The eye at P′ then sees the effective aperture of the prism traversed by fringes which, when broadened as much as possible, tell us how a plane wave-front has been affected by the double passage through the prism.

These defects of the wave-front may either be due to imperfect planeness of the surfaces of the prism or to imperfect homogeneity of the glass, or to both causes. But to whatever cause they may be due, they can be corrected by local polishing of the surfaces. For instance, if the fringes show that the optical path through any part of the prism is longer than at others, by polishing a little hollow in the corresponding part of the surface the wave can be flattened out. In this way, by local polishing at the points indicated by the contour fringes, the prism can be made to transmit a perfectly plane wave-front. But it should be noted that the correction only holds when the light traverses the prism in the same way. The interference tests should always be made, therefore, with the light passing through the prism in the direction which it will pass in use. Thus for a spectroscope prism the prism should be adjusted for minimum deviation.

In applying the interferometer to instruments of the lens type which have not infinite focal length, the arrangement employed is that of *Fig.* 3 (*c*).

The lens L is mounted in a suitable chuck, with its axis parallel to the incident beam. It causes the beam to converge towards its focus at F. A convex mirror M_2' is placed in the position shown, and adjusted so that its centre of curvature coincides with F. The converging rays then meet the surface of M_2' normally, and, after repassing through L, form a parallel beam once more, and are focussed at P′ by L_2 as usual. If the lens were free from spherical aberration, so that rays from all zones converged to one focus, it would be possible to adjust M_2' so that the aperture of the lens as seen by the eye at P′ would be free from fringes. In the presence of aberration, however, the different zones have different foci, and in whatever position M_2' may be there will be circular fringes in some zone, indicating the amount of the aberration. In the event of the lens being astigmatic, the fringes will be elliptical.

This method has been extended by Twyman to quite complicated cases, such as photographic lenses at oblique incidence, microscope objectives, etc. As applied to lenses, however, the interference method must be regarded as still in the course of development; much has yet to be done in designing the best mechanical means of realising the necessary conditions of adjustment in order that the resulting fringe system will give the information which it is desired to obtain. There is, however, little doubt that the use of interferometer methods for the examination of almost every kind of special instrument will ultimately be widely adopted.

In employing the interferometer for such purposes it is necessary, as with the arrangements of *Fig.* 1 ((*b*) and (*c*)), to obtain exact collimation and normal incidence of the beam on the equivalent film.[1]

J. G.

INTERMITTENT LIGHTS, PHOTOMETRY OF. See "Photometry and Illumination," § (122).

INTERNATIONAL CANDLE: the unit of candle-power adopted in 1921 by the International Commission on Illumination. See "Photometry and Illumination," § (14).

INTERRUPTER: an apparatus for mechanically interrupting the primary current of an induction coil. See "Radiology," § (16).

[1] Guild, "Fringe Systems in Uncompensated Interferometers," *Proc. Phys. Soc.*, 1920, xxxiii. 40.

Interval: a term used in music to denote the musical relation between two sounds. See "Sound," § (2).

Intervals, Chief, within an Octave. See "Sound," § (6) (vi.).

For various temperaments, tabulated. See *ibid.* § (6) (vi.), Table II.

Intrinsic Brightness: a term used to denote the measure of the light-emissive power of a surface per unit of area of surface; generally quoted in candle-power per square inch. See "Projection Apparatus," § (2).

Invar: its applications to tapes and wires used for base measurements. See "Surveying and Surveying Instruments," § (39). See also "Invar and Elinvar," Vol. V.

Iodeosin: a reagent for testing glass surfaces. See "Glass, Chemical Decomposition of," § (3) (i.).

Iris: the diaphragm which limits the aperture of the eye. See "Eye," § (2).

Irradiation: the term in radiometry which corresponds with illumination in photometry. See "Spectrophotometry," § (15).

Iso-lux Diagram: a diagram of equal illuminations, analogous to an isobar diagram. See "Photometry and Illumination," § (69).

J

Jäderin, E. His method of using tapes and wires in catenary. See "Surveying and Surveying Instruments," § (39).

Jellets' Prism: a polarimeter which depends on the photometric principle of matching two illuminated fields by varying their relative intensity. See "Polarimetry," § (3).

Joly Photometer. See "Photometry and Illumination," § (29).

K

K Series: a group of spectrum lines in the characteristic X-rays emitted by an element. See "Radiology," § (17).

KINEMATOGRAPH

§ (1) Introduction.—Although the name kinematograph has been selected for the title of this section as being more in accordance with the Greek derivation, other terms such as cinematograph, or more briefly cinema, the first syllables of which are pronounced softly, have received greater public recognition. Over sixty designations have been introduced from time to time, but, with the exception of a few proprietary names, practically all have now lapsed in favour of those mentioned above.

The history[1] of the highly perfect apparatus of the present day is a record of comparatively slow and intermittent progress during half a century, culminating in the introduction of the continuous photographic film by Friese Greene in England and its perfection by Edison in America, and, in consequence, a rapid development of the apparatus as a means of public entertainment and, to a more limited extent, of education.

§ (2) General Description.—The primary function of the kinematograph is to reproduce pictorially upon a screen the movements of objects. This is done by presenting to the eye in regular order a series of pictures each of which represents a consecutive stage in the motion, the speed of presentation being such that the eye appreciates the series as a continuous picture in which the objects may appear in motion. By means of a special camera[2] a series of instantaneous photographs of the moving object is taken, with exposures of about $\frac{1}{100}$ second, upon a sensitised transparent celluloid film which is advanced stage by stage in the focal plane of the objective at intervals of about $\frac{1}{16}$ to $\frac{1}{20}$ second. During the stationary periods the film is exposed and during the transitions the light is occulted by the shutter. Thus the series of photographs does not comprise the complete movement of the object. Short alternate stages are unrepresented.

When the images are projected upon the screen by means of the kinematograph projector, a true representation of the original movement is obtained when the speed of the movement of the operation and the ratio of the bright and dark intervals are the same as those pertaining during the taking of the series of photographs with the camera, but in actual practice there are small departures from true reproduction arising from differences in the velocity ratios.

Although the image projected upon the screen may appear to be in continuous motion, the action is actually discontinuous. A stationary image representing one stage remains on the screen for a period of about $\frac{1}{20}$ second. This is followed by a dark interval lasting about $\frac{1}{80}$ second. Although the original

[1] *Living Pictures*, Hopwood and Foster, 1915.

[2] "Motion Picture Cameras," C. L. Gregory, *Trans. Soc. Motion Picture Engineers*, April 1917.

stages of the movements corresponding with these brief dark intervals are not reproduced upon the screen, the mind of the observer fails to detect their absence, and the interrupted series of images when combined by the retino-cerebral apparatus is accepted as being equivalent to the original continuous appearance.

To obtain this stroboscopic effect it is essential that the dark intervals should be so small that the impression on the mind of one stage has not become unduly weak before the succeeding impression has been formed. That the effect is entirely due to a physiological process of fusion of the successive images has been contested [1] on the experimental grounds that under certain conditions an impression of movement is obtained even when the dark intervals are also distinctly appreciated by the mind.

If the visual sensation ceased immediately the direct excitation had ceased, that is, if the phenomenon of persistence of vision did not exist, one image sensation could not extend over the dark interval and overlap, as it were, the succeeding image. Under these circumstances the complete synthesis of a series of images separated by brief dark intervals would hardly be possible, and kinematography would be impracticable.

§ (3) The Eye and Vision — Flicker.— The functions of the eye as the organ of sight are discussed in the article on "The Eye and Vision" (*q.v.*).

In its dark adapted state, as under the conditions of a modern picture house, the retina can be subjected to a range of intensity of from one to about a thousand without producing a glare effect. But in practice, under these conditions, a sudden change of intensity from one to more than a hundred or two should be avoided to ensure entire absence of visual discomfort. If the eye were repeatedly turned from the illuminated screen to very dark surroundings, the contrast might easily be such as to cause discomfort in the course of time, and therefore a certain amount of general illumination, which also facilitates the entrance of the spectators, is customary, the contrast being kept, however, within the above-mentioned limits. As the illumination of the picture on the screen varies from about 1 to 1·5 foot-candles in the case of the quarter lights to about 5 or 6 foot-candles over the brightest parts, a general illumination of 0·07 foot-candle is permissible. Much, however, depends upon the circumstances, and general illumination of twice this amount has been advocated.[2]

Reference has already been made to the persistence of vision upon which kinematography essentially depends. The first definite measurements of the time of persistence are those recorded by M. D'Arcy,[3] who measured the time of whirling necessary to produce the appearance of a continuous circle of light when a live coal was whirled at a distance of 165 feet from the observer.

When the retina is excited by a single low-intensity impulse of very short duration, the resulting pulse of sensation takes an appreciable time to grow and a similar time to wane, the duration of the sensation being longer than that of the stimulation. According to McDougall, a stimulus of greater intensity produces a series of partially overlapping pulses of rapidly diminishing maximum intensities.[4] Charpentier's bands, which may be observed when a suitable radial slot is rapidly rotated, are attributable to pulses of this kind. The initial pulses of greatest maximum intensity wax and wane rapidly. Succeeding pulses do so more slowly, and within the fraction of a second the intensity of the pulses becomes too small to be appreciated. Recurrent images and other phenomena of vision [5] that involve periods much greater than those of the kinematograph need not be considered here.

Flicker, the elimination of which is a problem of such great importance in the intermittent type of kinematograph, is dependent upon the duration of the individual impulses that fall upon the retina. During the period of transference of the consecutive pictures it is necessary to interrupt the projection of the image upon the screen. This is done almost universally by means of a rotating shutter having a blank sector which intersects the beam of light between the objective and the screen during the period of the movement. Thus the illumination of the screen is intermittent and the retina is subjected to a rapid series of impulses which, under certain conditions, produce a disturbing appearance of flicker, that may even prove injurious to the eyesight if long continued.

Flicker can be eliminated by increasing the frequency of the interruptions or by reducing the illumination. It is also dependent to some extent upon the relative durations of the consecutive black and white periods,[6] the maximum effect being obtained when the white interval equals the black. Thus suppose the disc is half black and half white and the speed such that flicker is pronounced. If then the white sector is increased at the expense of the black, the flicker will diminish and be entirely absent when the disc is all white,

[1] P. Linke, *Die stroboskopischen Erscheinungen als Täuschungen des Identitätsbewusstseins*, Leipzig, W. Engelmann, 1907.

[2] L. A. Jones, "The Interior Illumination of the Motion Picture Theatre," *Trans. Soc. of Motion Picture Engineers*, No. 10, May 1920.

[3] *Mémoires de l'Académie des Sciences à Paris*, 1765, p. 450.

[4] *British Journal of Psychology*, 1904, i. 78.

[5] Bidwell, *Curiosities of Light and Sight*, 1899, chap. v.

[6] Marbe, *Theorie der kinematographischen Projectionen*, sec. 10, p. 43.

since there are then no interruptions of the light. Similarly, if the black is extended at the expense of the white, the flicker will progressively be diminished and again vanish when the disc is all black. In the kinematograph the black sector which cuts off the light during the transference of the picture often covers about 90° of arc, that is, 25 per cent of the light is cut off on this account. The remaining 75 per cent of the light would be available for the illumination of the screen if the speed was sufficiently high to eliminate flicker. As sixteen pictures are projected per second, the shutter makes 16 revolutions per second, which in practice is nearly three times too slow so far as the elimination of flicker is concerned. It is necessary, therefore, to increase the number of interruptions per second by the introduction

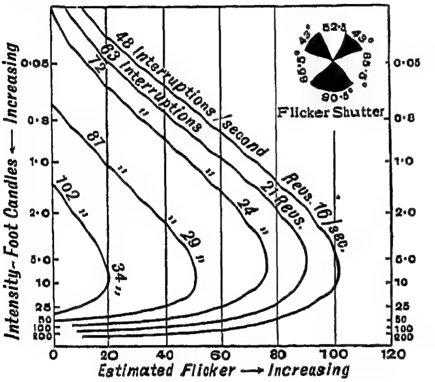


FIG. 1.

of one or more, usually two, additional dark sectors, the arcs of which are generally made less than that of the main occulting sector, in order to conserve the illumination.

Results of flicker tests of a three-bladed shutter supplied with one well-known type of machine are indicated in *Fig.* 1. In the tests the film itself was removed and the fluctuations of the light in the aperture were measured photometrically. Abscissae represent the estimated strength of flicker and ordinates the intensity of illumination in foot-candles. Upon each curve is indicated the corresponding number of interruptions per second, that is, three times the number of revolutions per second of the particular shutter. Thus, for example, in the case of the curve corresponding with a speed of 29 revolutions per second, the interruptions of the light were 87 per second. When the intensity of the illumination was reduced to 0·7 foot-candle the flicker disappeared. As the intensity was increased flicker reappeared and became more and more pronounced until it reached a maximum value at an intensity of 10 foot-candles. A further increase of the intensity produced an appearance of glare and caused the flicker to disappear very rapidly. Glare effects of this order need not be considered as they do not occur in the case of the kinematograph.

The curves indicate that with an intensity of illumination of 2 foot-candles about 36 revolutions per second would be required to eliminate flicker. In practice, however, this condition is fortunately attained at a much

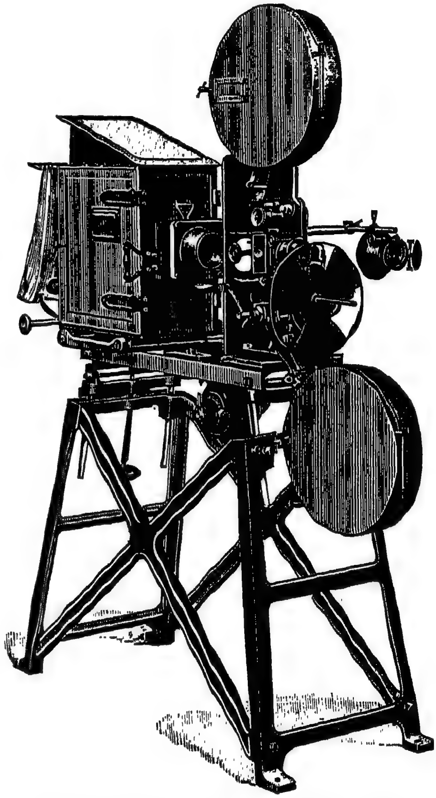


FIG. 2.—The "Indomitable" Projector Outfit.

lower shutter speed of 16 revolutions per second, equivalent to 48 interruptions in the case of the three-bladed shutter. This is largely due to the fact that with the film in position not only is the illumination reduced by at least 25 per cent in the case of the lightest portions, but also there is no longer a rapid change from uniform light to darkness, since the light is broken up by the various tones of the picture, thus further reducing the general illumination and contrast with the dark intervals.

§ (4) THE APPARATUS.—A kinematograph projector equipment, of which the Indomitable outfit (*Fig.* 2) of Messrs. Kershaw & Sons is a

typical example, comprises the lamp, lamp-house and condenser, the machine or projector, including the projector lens, and the stand. The primary function of the machine is to draw the film from the upper reel box; to bring each consecutive picture of the film very exactly into position before the gate aperture through which the light passes; to keep the picture stationary in this position for a period of about $\frac{1}{20}$th second; to remove the picture and bring the succeeding one into position in the brief interval of about $\frac{1}{80}$th second; to wind the film upon the lower reel after it has passed through the gate before the aperture; to interrupt the image falling upon the screen, particularly during the transference period; and to mask the picture, that is, to adjust the film so that the individual pictures lie within the boundaries of the aperture and thus lie correctly on the screen.

FIG. 3.

In order that the film may be fed through the machine regularly, the sides are perforated in the manner indicated in *Fig.* 3, the perforations in question being now standardised.[1] Similarly, the sprocket wheels or drums have their teeth, which engage the film perforations, also pitched exactly according to a standard. Upon the regularity and accuracy of the perforations and the teeth depends to a large extent the steadiness of the picture upon the screen.

In the diagram, *Fig.* 4, the film A is drawn from the upper reel box B by means of the uniformly driven upper feed sprocket C, the teeth of which engage the film perforations. A roller D holds the film in engagement with C, from whence it passes through the gate E in front of the aperture F through which the light from the condenser passes. Friction strips, which press upon the sides of the film and not upon the part occupied by the pictures, offer a certain amount of resistance

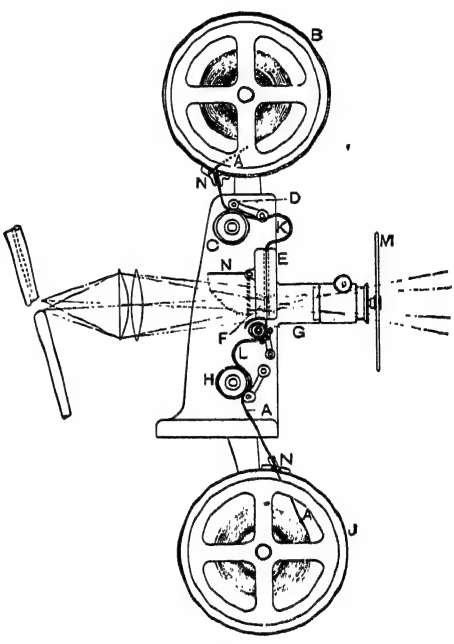


FIG. 4.

to the passage of the film as it is pulled through the gate by means of the intermittent sprocket G. From G the film passes over the continuously driven lower feed sprocket H into the lower reel box J, the spindle of which is frictionally driven in order that the rate of winding may be constant notwithstanding variation of the reel diameter. As the feed sprockets C and H rotate uniformly while G rotates intermittently, it is essential to provide loops at K and L the lengths of which must not be less than one picture, as otherwise the film would be broken. If the loops, on the other hand, are too long an objectionable whipping noise may result.

[1] *Report and Recommendations of the Standards Sub-Committee, Incorporated Association of Kinematograph Manufacturers, Ltd.*

In certain machines in which masking is performed by a displacement of the intermittent sprocket, the lower loop must be sufficiently long to permit of masking to the extent of at least one complete picture.

§ (5) The Intermittent Feed Mechanism.—Of the elements above described the most interesting and indeed the most important is the intermittent feed mechanism, because upon the accuracy with which it places each picture in front of the gate aperture depends the steadiness of the picture upon the screen. Assuming that a variation in the position on the screen of consecutive pictures of $\frac{1}{4}$ inch is permissible, and that the distance of the objective from the screen is 100 feet and from the film 5 inches, then the pictures must be centred with an accuracy of $\frac{1}{1000}$ inch.

Innumerable devices have been proposed for the intermittent movement of the film, but in the great majority of projectors the Maltese cross arrangement, or the Geneva cross, as it is sometimes called, represented in *Fig.* 5, is employed. Upon the axis E of the gate sprocket which feeds the film there is mounted the cross H, having four radial slots and four concave sides which suggest the name accorded to this mechanism. On a parallel axis F there is mounted a disc and driving pin A, the disc being cut away as indicated at K (*Fig.* 6) in order to clear the extreme points of the cross. The drive is applied to this driving disc, which rotates uniformly in one direction only.

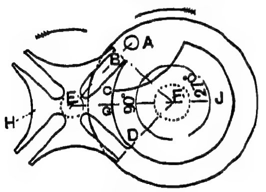


Fig. 5.

In the relative positions indicated in *Fig.* 5, the driving pin A is approaching the slot B tangentially, and the Maltese cross, the gate sprocket associated with it, and the film are at rest. Rotation of the cross is prevented in one direction by the locking action of the disc portion GD on one side of the line joining the centres EF and in the other direction by the portion CG. As the pin approaches the slot this locking portion decreases, but the drag on the film tends to hold the cross against the other locking portion.

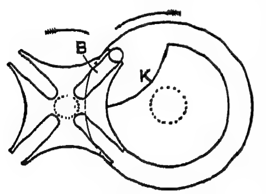


Fig. 6.

In the position *Fig.* 6, the pin has just entered the slot B and is driving the cross in the direction of the arrow, and with it the film. The recessed portion K of the disc is necessary to clear the points of the cross as indicated in *Fig.* 7.

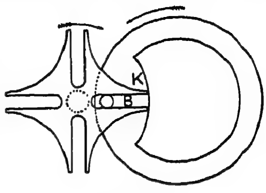


Fig. 7.

In *Fig.* 8 the driving pin A has just left the slot B of the cross, which with the film has therefore just come to rest, and the cross is locked to the maximum extent by the disc. As the cross has four symmetrical slots it makes a quarter revolution for one revolution of the driving disc.

Further, as the angle BFD (*Fig.* 6) is 90°, the cross during one rotation of the driving disc is being driven for one-quarter of the period and is at rest during the remaining three-quarters. The gear is said to be three to one, since the picture on the film rests before the gate aperture during three-quarters of the period and is transported during the remaining fourth.

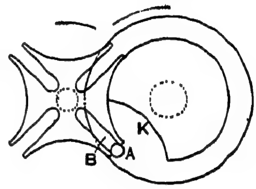


Fig. 8.

Owing to the occultation of the light during this period only three-quarters of the total light reaches the screen, and in practice considerably less, owing to the additional width of the black sector required to cover nearly the width of the objective and to the flicker sectors previously referred to. As illumination is of great importance the width of the occulting sector is reduced by the adoption of a four-to-one, or even a five-to-one gear, as indicated in *Fig.* 9. The driving pin engages the cross during 60°, or $\frac{1}{6}$th of its rotation, and leaves it at rest during the $\frac{5}{6}$ths of the total period. The occulting sector can thus be reduced from about 90° to about 60°, with a corresponding increase of illumination. But

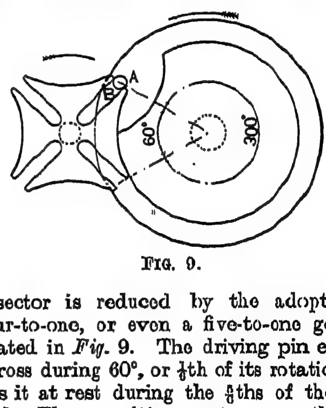


Fig. 9.

it will be observed that the pin A, instead of gently entering the slot B tangentially, forcibly strikes the driving face, and as the film has to be started from rest, moved, and brought to rest again in $\frac{1}{6}$th instead of $\frac{1}{4}$th of the period, the strain on the film, and particularly the wear of the perforation, is much greater. Notwithstanding these disadvantages four- and five-to-one arrangements are frequently adopted.

A greater strain is thrown on the film at the moment of starting than of stopping, as the film has to be drawn through the gate against the resistance of the friction side blocks. An excentric arrangement [1] of the slots as indicated in *Fig.* 10 has the advantage

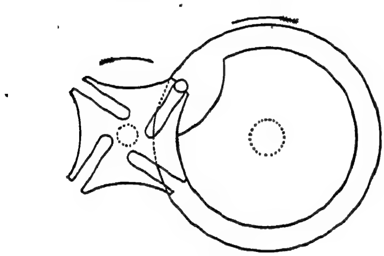


FIG. 10.

that, although the action may still be, say, five to one, the film is started more slowly, the movement being decelerated correspondingly more rapidly; but, as already stated, rapid starting is more serious as regards wear than rapid stopping. For the excentric arrangement the further advantage is claimed that a portion of the occulting sector on the leading side can be cut out with a consequent increase of the light, since a greater part of the initial movement of the picture is too slow to be recognised as a ghost image by the eye; but it should be remembered that it may be necessary to extend the following side, which under ordinary circumstances is often reduced in practice below the theoretical amount.

Claw-feed mechanisms of many kinds have been introduced, but although they are frequently adopted in camera, perforating, and printing mechanisms, they are but rarely used in the projector. Whereas in the projector the picture rests for $\frac{1}{20}$ second and moves during $\frac{1}{80}$ second, these periods are more or less reversed in the case of the camera, as the time of exposure during which the film is at rest is less than $\frac{1}{100}$ second.

For printing machines the claw arrangement [2] has the advantage that the teeth engage simultaneously the perforations of the positive and the superposed negative and drive both together with great precision. A considerable wear of the film perforations is attributable to the teeth of the claws, but as the film is only subjected once in the camera to the claw action, this disadvantage is not of great importance. The teeth of the claw are caused to move in a more or less D-shaped path, the straight side of the D being parallel to the film. The vertical motions are controlled by a main cam and the horizontal movements by a subsidiary cam.

One of the earliest devices, known as the Dog or Beater, is now only found in machines of the cheapest kind. While simplicity of construction is its chief characteristic, its destructive action on the film is considerable. The device was originally suggested by the appearance of a lathe dog, from which the name is derived. At each revolution of the dog the pin strikes the film and beats it downwards, thus drawing the picture through the gate. The throw of the dog is such that the film is displaced to the extent of one picture at each revolution. A cam is usually provided to press the gate open against the resistance of a spring immediately before the striking pin comes into action in order to reduce the strain on the film. A reciprocating arm is sometimes used instead of the rotating dog, not only as a feed device but, in some cases, as a means of masking. Particulars of numerous other devices will be found in works devoted to kinematograph details.[3]

§ (6) THE SHUTTER.—With very few exceptions, the shutter M (*Fig.* 4), placed immediately outside the objective, has the form of a rotating disc which intersects the beam of light. It is driven in direct association with the driving disc of the Maltese cross, one revolution of which corresponds with the interval between the pictures. Thus, when properly set, the dark sector of the shutter occults the light while the picture is being transferred. The angular width of the occulting sector should theoretically be such that it just covers the whole beam before the picture commences to move, and does not uncover any portion of it before the picture has come to rest. To satisfy this condition would involve the loss of a great part of the light, especially as flicker sectors are also required. If the transfer takes place during $\frac{1}{5}$th of the period—that is, during 72° of rotation—the angular width of the sector is generally between 90° and 95°, which usually means that some light passes at the commencement and end of the transfer, when, however, the motion of the image is comparatively slow. That some latitude is permissible is suggested by the fact that in some houses where the film is run at an unusually high speed,[4] or where the illumination is low, due to fog or smoke, for example, a special single-bladed shutter, or sometimes no shutter, is used.

[1] W. B. Cook, "The Excentric Star Intermittent Movement," *Trans. Soc. of Motion Picture Engrs.*, May 1920.

[2] Dr. Forch, *Der Kinematograph*, p. 30.

[3] Hopwood and Foster, *Living Pictures*, 1915; Richardson, *Motion Picture Handbook*, 3rd ed., 1915; Forch, *Der Kinematograph*, 1913; Liesegang, *Handbuch der praktischen Kinematographie*, 6th ed., 1919; Lehmann, *Die Kinematographie*, 1919; Lassally, *Bild und Film*, 1919, vols. i. and ii.

[4] F. H. Richardson, "The Various Effects of Overspeeding Projection," *Trans. Soc. Motion Picture Engineers*, No. 10, p. 61.

Several typical shutters in common use are illustrated in *Fig.* 11.

Innumerable devices have been introduced with the object of improving the illumination, but the result must necessarily be a compromise between light and definition, as practically all involve the projection of diffuse or irregular light upon the screen. Thus the blank sectors may be pierced with holes, have gauze-covered portions, be of semi-transparent, variegated, or coloured material, or be provided with a small-diameter, light-diffusing lens.

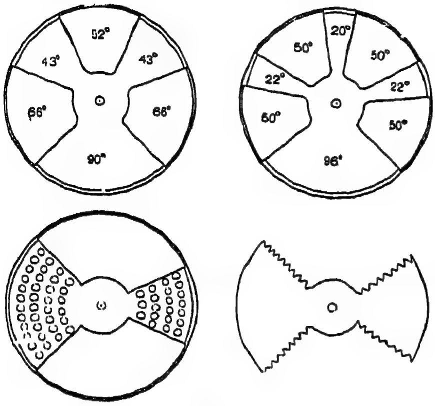


FIG. 11.

§ (7) THE GATE.—An important element is the gate (*Fig.* 12) through which the film is drawn downwards intermittently by the intermittent sprocket. The film passes through the space B between the aperture plate A and the gate E, which is so hinged at C that it can be quickly opened by pressing the spring catch D to facilitate the stringing of the film.

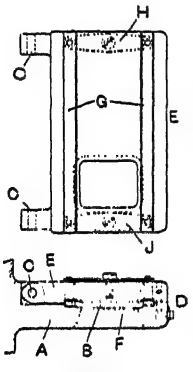


FIG. 12.

Usually, but not in all cases, the aperture plate faces the lantern and the gate the objective. When the gate is on the objective side it is often necessary to swing the objective holder upwards before the gate can be opened to the full extent. When the gate faces the lantern, the disposition of the guide rollers is usually such that if the film breaks it accumulates between the machine and the lantern and increases the danger of fire. Inside the aperture face there are provided two side strips, against which the film is pressed by two corresponding hard steel strips or shoes G, carried springily upon the gate. As it is the pressure of these shoes that determines the drag on the film, it is essential that not only should the tension be capable of adjustment, but that the tensions on the two sides of the film should be equal. This result is usually attained by making the top tension spring H act equally upon the upper ends of the two shoes and the lower spring J upon the lower ends. In some cases one spring is arranged to act equally on all four ends.

Small differences in the width of the film necessitate some side guidance, as the width of the slot in the aperture plate must be sufficiently great to accommodate films of extreme width. The flange of the top guide roller may be made to serve as a side guide, as the use of springs, especially in the side wall of the aperture plate, is not free from objection on account of their tendency to become clogged with the collodion torn from the surface, particularly of new films. For the same reason the use of velvet strips has been practically abandoned.

To facilitate the cleansing of the guides the gate should be capable of being opened widely and the shoes themselves should be removable. In many machines the lower guide roller which holds the film against the intermittent sprocket is carried upon the gate, and is capable of being so adjusted that it does not bear too hard upon the film. Underneath the sprocket there is mounted a stripper plate which separates the film from the under side of the sprocket. When this plate is not fitted there is danger of the film being wound round the sprocket.

§ (8) MASKING.—Masking the film is necessary from time to time, in order to set the individual pictures exactly before the aperture and therefore correctly upon the screen.

In re-joining a film the pieces may be displaced by a fraction of a picture, and when this portion passes through the machine the correct location of the image upon the screen is affected. The earlier method of masking, which has only recently been generally discarded, consisted in displacing, relatively to the film, the optical axis containing the source, the centre of the aperture, and the optical centre of the objective, by the amount required to replace the image centrally upon the screen. The model illustrated in *Fig.* 2 is of this type. The aperture and objective are usually moved together, and then the light source is readjusted until the best condition of uniform illumination is again attained.

As the frequent readjustment of the source is not without objection, the greater proportion of modern machines have the optical axis fixed, masking or framing being effected by pulling the film through the gate until the picture is again central, without, however, altering the relative positions of the Maltese

cross and the driving disc with which the shutter is associated; that is, the operation of masking must not affect the timing of the carry-over gear and shutter. The intermittent sprocket, Maltese cross, and driving disc are moved up or down as one unit carrying with it the film relatively to the gate. It is then necessary to make the top and bottom loops sufficiently large to permit of this adjustment. The drive may be communicated to the adjustable element through the intermediary of a coupling so disposed that no rotation of the coupling shaft and no rotation of the shutter shaft is involved in the masking operation. If these conditions are fulfilled, the timing is not affected, the occultation of the light corresponding with the transfer of the film by the rotation of the intermittent sprocket.

This system of masking is embodied in the Power's Cameragraph. The apparatus, as is customary, can be placed in an inclined position, which is necessary when the projection booth is situated at a high level relatively to the screen. Means are provided on the stand for angling the whole machine together with the lamp as one unit to suit the conditions of the installation.

An interesting device is used in the Cameragraph for automatically re-setting the lower loop, which in practice may quite easily be lost if the film perforations, for instance, are badly worn, necessitating ordinarily an interruption of the projection. The strip of film forming the lower loop is passed under a roller. When for any reason the loop is lost, the film in becoming taut raises the roller and partially rotates a cylinder on the surface of which is cut a spiral slot. An arm, a pin on which engages the slot, is thus swung sideways and declutches the lower feed sprocket, the rotation of which momentarily ceases. During this interval the loop increases, the roller under the action of a spring falls with the loop, the clutch is brought into normal action, and the sprocket recommences to feed the film.

In the masking devices previously described the intermittent sprocket in its lowest position is a considerable distance from the gate, and the length of the film loop between the gate and the intermittent sprocket is of variable amount. Although additional complexity is involved, this disadvantage may be successfully avoided by mechanism such as is used, for example, in the Simplex machine. The gate sprocket is mounted close to the gate with its axis in a fixed position. Masking is effected by the rotation of the whole intermittent mechanism about the axis of the sprocket, the maximum rotation required being a quarter revolution, which corresponds with a film displacement of one picture. Since the pinion of the driving disc must necessarily be in gear either directly or indirectly with the driving wheel, having its axis concentric with the intermittent sprocket about the axis of which the masking rotation takes place, it will be evident that in the masking operation the driving disc will be rotated relatively to the Maltese cross. To preserve the timing, the shutter must therefore be automatically rotated by an equivalent amount in the appropriate direction.

For a displacement of one picture—that is, a quarter revolution of the intermittent sprocket—it is necessary to correct automatically the displacement of the shutter by one revolution. This is done in one particular example by communicating the drive to the shutter axis through spiral gears, the pinion gear on the shutter axis having a length of one spiral. When the gate sprocket mechanism is rotated the spiral driving element is displaced longitudinally upon a squared shaft by a corresponding amount, and in its longitudinal movement it rotates the shutter quite independently of its ordinary rotational movement. The arrangement has the advantage that, by providing an additional hand control of the longitudinal movement, the timing can be adjusted while the machine is running.

In view of the precision with which the intermittent sprocket must locate the film, the introduction of elements between this sprocket and the Maltese cross is objectionable owing to the possibility of backlash introduced by slackness or wear of the parts. It has been proposed, for example, to rotate the sprocket independently of the Maltese cross through the intermediary of a differential, one element of which is controllable by hand. In another arrangement the sprocket is mounted upon a spiral sleeve on the Maltese cross spindle to which it is keyed. By a longitudinal displacement of the sleeve the sprocket can be rotated relatively to the cross, and, as in the previous example, the timing of the shutter is not affected, since no relative movement of the cross and driving disc is involved. Devices of these kinds, which involve the use of additional parts, particularly between the sprocket and the cross, are unlikely to prove satisfactory on account of the objectionable backlash that may be introduced.

An essential and important feature is the take-up mechanism of the lower spool-box which rolls up the film after its passage from the lower feed sprocket at a constant rate of about 1 foot per second. As the diameter of the roll increases the speed of rotation necessarily decreases and the tension on the film varies. In a few cases the drive is applied through the intermediary of a variable speed gear, but in the great majority of machines a simple slipping clutch is employed, the pressure being adjustable by means of a spring. As the diameter of the roll grows larger the slipping increases, but although the maximum force that can be applied to the film is limited, the force is not automatically kept constant. For reels having 1000 feet of film the slipping-clutch, when properly set, suffices, but for longer reels of 2000 feet a variable speed drive becomes desirable. Satisfactory results have been attained by causing the increasing weight of the roll to increase the frictional resistance to rotation, whereby the uniformity of the tension on the film is maintained.

So long as inflammable celluloid film is used, there must always be a certain amount of fire risk, notwithstanding the stringent legal regulations that are commonly imposed. For the prevention of panic among the spectators, who may be alarmed by the fiery glow projected

upon the screen when the film in the aperture space is ignited, it is essential that the projection should be interrupted immediately by closing the window shutter of the booth and by closing the lamp shutter or dowser. These operations are generally done by hand. Upon the machine itself, however, there must be provided a fire shutter N (*Fig.* 4), which covers the gate aperture when the machine is at rest, as shown dotted in the illustration.

When the machine is in operation the film is subjected to the concentrated heat for brief intervals of only $\frac{1}{16}$th second, during which time it does not reach the temperature of ignition. But if the machine is stopped and the film is at rest for a second or even less, it will become ignited unless the shutter falls immediately the motion ceases. The shutter is often operated by means of a governor directly associated with the driving mechanism, and in other cases by simple friction discs separated by a thin viscous layer of oil.

If the film breaks while the machine continues in motion the fire shutter will not fall. In such a case the lamp dowser must be closed by hand.

To provide automatic protection capable of dealing with every contingency involves mechanical and electrical complications that may prove a greater danger than security. Reliance is usually placed upon the fire shutter at the gate and the alertness of the operator. Ignited film in the machine itself, being in contact with cold metal, burns comparatively slowly and may easily be extinguished if dealt with promptly.

It is most important that the large masses of film in the roll boxes should not become ignited. Provided the lids are closed as they ought to be, this can be effectively prevented by passing the film into the boxes between metal plates N (*Fig.* 4), having a surface length of several inches. The cold metal extinguishes any flaming film that may be drawn between them.

§ (9) The Objective.—As the individual pictures on the film are only 1 inch wide and $\frac{3}{4}$ inch high, whereas the image projected on the screen may be 20 ft. wide, it will be evident that so great a magnification necessitates a quality of definition of a very high order over as large a central area as possible, since the action of a picture play cannot be confined to the centre of the screen. Illumination being of the greatest importance, a large aperture is desirable as in the case of portrait lenses, in which the conditions to be satisfied are very similar. There may be a considerable loss of light, however, in the passage through the lens system, not only at each transmission, but also owing to the actual cutting off in the combination when the front and back elements are widely separated as in the case of the early Petzval objectives. Compactness of design is therefore of importance.

An objective equivalent focal length of 4 inches is a common one, and since the picture is 1 inch wide the angular field is about 15°, and for longer focus objectives still less. This small angular field makes it possible to obtain a quality of general definition on the screen that would be unattainable if the width of field common in the case of a portrait objective was required.

Computation is also facilitated by the fact that complete freedom from distortion is not of greatest importance. With very few exceptions, the centre of the projected beam does not fall normally upon the screen, owing to the elevated position of the kinematograph booth and to the vertical arrangement of the screen. For these reasons the magnification at the bottom of the screen is usually greater than at the top by an appreciable amount considerably in excess of the distortion error attributable to a good objective.

The principal aberrations that have to be reduced to a minimum in the computation of objectives are discussed in the undernoted works.[1] They are as follows:

(i.) Chromatic aberrations.
(ii.) Axial spherical aberration.
(iii.) Coma.
(iv.) Astigmatism and curvature of the field.
(v.) Distortion.

Freedom from chromatic aberrations ensures the absence of coloured margins in the projection of black and white details and a true representation of coloured images. Crispness of central definition is associated with freedom from axial spherical aberration, and fulfilment of the sine condition. Coma or spherical aberration for extra axial points and astigmatism result in fuzziness and elongation of details which increase with the distance from the centre.

When there is curvature of the field the image cannot be sharply focussed at the centre and sides of the screen simultaneously, owing to the image surface being curved. When astigmatism is present there are really two image surfaces whose vertexes coincide on the axis where the astigmatic aberration is zero. Distortion, already referred to as being of secondary importance in view of the greater error due to the angle of projection, results from a change of magnification from the centre radially over the surface of the screen.

Most kinematograph objectives are based upon the type of portrait objective invented by Professor Petzval of Vienna in 1840. This type is remarkable for its freedom from aberration over a central area corresponding with an angular field of about 8°, that is, 4° on either side of the axis. Beyond this width non-fulfilment of Abbe's sine condition becomes increasingly marked.

[1] Steinheil and Voit, *Handbook of Applied Optics*, edited by J. W. French, vols. i. and ii., 1918; A. S. Cory, "Optical Requirements of Motion Picture Projection Objectives," *Trans. Soc. Mot. Pict. Engrs.*, April 1918; C. Lindsay Johnson, *Photographic Optics*.

The Petzval system comprises an outer cemented doublet nearest the screen and an inner non-cemented doublet nearest the film. It has the additional advantage that the latter is not so liable to be affected by the heat of the concentrated light as would be the case if the elements were balsamed. In 1866 J. H. Dallmeyer considerably improved the uncemented element of the Petzval system.

By a judicious selection of types of glass introduced within comparatively recent times it is now possible to obtain numerous excellent anastigmatic combinations of British, German, and French make that, in addition to crispness of definition over a large area, provide a field that is flat over nearly the whole angular width common to kinematograph projection.

§ (10) Illumination. — When it is considered that the whole light available for distribution over the surface of the screen, which may be 20 feet long and 15 feet high, has to be emitted by the crater of the arc having an effective surface of less than 0·2 sq. in., it will be understood that a very intense source is essential. For smaller projection lengths special incandescent tungsten lamps [1] having a condensed area of filament are extensively employed in America. With these lamps there are used rear reflectors and special corrugated condensers [2] which, being placed near the lamp, intercept a large cone of rays. But for screens of 20 feet width a more powerful source, such as the carbon arc, is still employed.

In comparison with the incandescent lamp the carbon arc [3] is more inconvenient to use, less constant, and less hygienic. If the carbons are solid the arc wanders and the illumination for this reason varies greatly, but usually the upper carbon has a soft core which, by burning away more quickly, centres the arc and prevents wandering. When using alternating current both carbons are usually cored. Arc controllers designed to avoid hand control, and magnetic devices for steadying the arc, are available and are sometimes installed.

When using cored carbons an illumination of about 120 candle-power per sq. mm. of crater is obtainable. An effective crater area of 100 sq. mm. or 0·16 sq. in., which is common, thus emits 12,000 c.p. If the arc was an unobstructed point source radiating in all directions, the nearer the condenser lens was placed to it the greater would be the cone of light intercepted, and the greater the light available. Much of the light, however, is obstructed by the lower negative carbon of direct-current arcs and by the upper crater edge of the positive carbon, and further, on account of the intense heat, it is not practicable to bring a condenser lens closer than 3½ in. to 4 in.

A 4½-in. diameter condenser lens, which is a usual size, set 3·6 in. from the arc intercepts a cone of light whose base covers 1 sq. ft. and whose length is 1 foot. Such a condenser thus intercepts 12,000 lumens if the candle-power is 12,000. At each transmission surface at least 4 per cent of the incident light is lost, in accordance with the Young-Fresnel law. There is also some absorption of light by the glass. In the case of the objective this is almost negligible, as the glass employed is of good optical quality and the thicknesses are small. Glass of a much inferior character is too frequently employed for the condenser lenses and the loss is then quite important. It is very questionable economy to use condenser lenses made of a type of window glass, as is so frequently the case.

The losses of light in its passage from the first condenser surface facing the arc to the screen are as indicated in *Fig.* 13. At the surfaces A, B, C, and D at least 4 per cent of the light incident at each of the respective surfaces is lost. An absorption of 5 per cent per cm. thickness of glass has been assumed, but this amount is often exceeded. In the first lens the loss is 17 per cent and in the double lens condenser 33 per cent. The loss in a triple condenser is still greater.

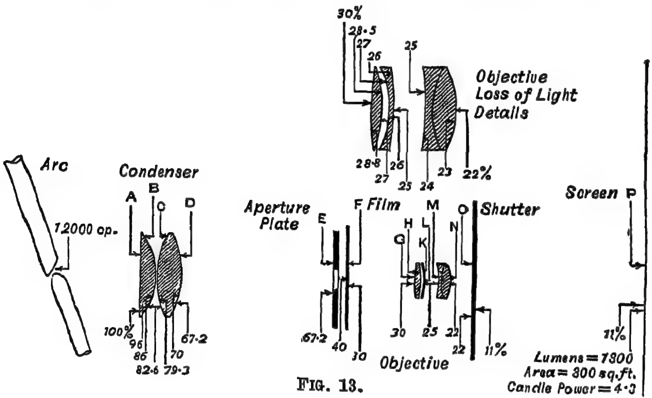


Fig. 13.

[1] R. A. Dennington, "Incandescent Lamps for Motion Picture Service," *Trans. Soc. Mot. Pict. Engrs.*, April 1918, p. 36, also 47.

[2] H. P. Gage, "Condenser Design and Screen Illumination," *Trans. Soc. Mot. Pict. Engrs.*, April 1919, p. 63.

[3] W. C. Kunzmann, "Carbon Arc for Motion Picture Projection," *Trans. Soc. Mot. Pict. Engrs.*, Nov. 1918, p. 20.

It has been assumed that the whole circular area of the condenser lens contributes light. Actually, however, only a rectangular area corresponding with the rectangular form of the film aperture is utilised. About 40 per cent of the circular beam is cut off by the rectangular aperture at E. Thus about 40 per cent of the original light reaches the film F. Only an approximate estimate of the loss at the film is possible, as the density varies greatly. Photometric measurements show that 25 per cent of the light may be lost in transmission through even the least dense parts.

The objective, whose surfaces are lettered G to N, usually has two uncemented lenses which involve a drop from 30 per cent to 25 per cent, and one cemented doublet which involves a further drop to 22 per cent.

In the usual flicker shutter O there is a very large loss of light, amounting in one commonly used type to nearly 50 per cent. Of the total light incident upon the first surface A of the condenser, namely 12,000 lumens, only 11 per cent passes the shutter, that is 1300 lumens. Neglecting the loss through absorption by the atmosphere, which may be considerable, this quantity of light is available for distribution over the screen P, having an area cf, say, 300 sq. ft. Thus, assuming the whole film is of the lightest shade, the intensity of the light on the screen would be 4·3 foot-candles. The actual intensity reflected to the observer's eye is considerably less, and depends not only upon the nature of the screen but also upon the obliquity of the observer's line of sight.

§ (11) The Screen.—A perfect reflector if used as a screen would reflect a high proportion of the light, but only in a definite direction determined by the angle of incidence. A matt surface, on the other hand, diffuses the light over a wide angle, within which the picture is visible, but the light reflected in any one direction is correspondingly less. Thus, although the albedo value, that is, the ratio of the reflected light to the incident light, for a polished glass sheet silvered on the back may be 90 per cent, such a screen would be quite unsuitable as compared with a matt surface of fine white plaster which, when clean, has an albedo value of about 80 per cent but spreads the reflected light over a wide angle.

White cloth well stretched reflects from 70 to 75 per cent of the light. Cartridge paper has even a higher albedo value, but such hygroscopic materials readily deteriorate. The choice of a screen surface [1] depends also upon the maximum obliquity from which it must be viewed.

[1] Jones and Fillius, "Reflection Characteristics of Projection Screens," *Trans. Soc. Mot. Pict. Engrs.*, Oct. 1920, p. 59.

Glass screens [2] ground on the face and silvered on the back confine the greater part of the light to an angle of about 30° and may have a good reflecting value of over 80 per cent when clean, but moisture on the ground surface has a serious effect upon the spreading power of the parts affected. Screens coated with silver, aluminium, or other metallic paints are often used. Special claims are sometimes made for yellow-tinted surfaces that reflect a greater proportion of yellow light.[3] Whatever the type of screen employed, it is essential that its surface should be kept clean and dry, as otherwise the reflecting power may be much reduced.

§ (12) Special Forms of Machine.—It is not possible within the limits of this brief account to deal with the more special applications of the kinematograph.

For the projection of synthetically coloured pictures Friese Green in 1899 proposed the use of a red, green, and blue rotating shutter in front of the objective. In the original kinemacolour method successive exposures of the film in the camera are made through three and sometimes two filters consecutively, and each set of three or two consecutive pictures is projected simultaneously, and in other arrangements alternately, upon the screen through corresponding filters. Numerous other arrangements have been proposed for the synthetic production of natural colours,[4] but, in view of the expense and complications involved, simple hand-tinted films [5] are more frequently used whenever coloured pictures are desired.

The kinematograph is of great scientific value for the analysis of very quick motions,[6] the pictures being taken at a rapid rate in the camera and projected upon the screen at the rate of about sixteen per second.

Many inventors have attempted to devise a satisfactory dissolving view machine, the advantages of which need no description, but no real practical success can be recorded. The intermittent machine, notwithstanding its inherent difficulties, still remains unsurpassed for the projection of pictures under picture-house conditions. Most of the continuous devices hitherto proposed have involved a large number of lenses, prisms, or mirrors difficult to adjust or maintain in adjustment individually with the accuracy essential for the projection of a picture at a

[2] Cardwell and Burrows, "Light Intensities for Motion Picture Projection," *Trans. Soc. Mot. Pict. Engrs.*, Oct. 1917.

[3] Richardson, *Motion Picture Handbook*, p. 166.

[4] Hopwood and Foster, *Living Pictures*, 1915, chap. vii.; Kelley, "Natural Colour Cinematography," *Trans. Soc. Mot. Pict. Engrs.*, Nov. 1918, p. 30; Carl Forch, *Der Kinematograph*, chap. x. p. 119.

[5] Blair, "The Tinting of Motion Picture Film," *Trans. Soc. Mot. Pict. Engrs.*, May 1920, p. 45.

[6] H. Lehmann, *Die Kinematographie*, sec. iv. p. 66.

distance of a hundred feet, which is quite a common practice.

Many attempts have also been made to project stereoscopic images [1] upon the screen, but with little real practical success. The use of coloured spectacles to view suitably coloured pictures is unlikely to become popular. An impression of depth can also be obtained from the parallax appearance resulting from a translation of the camera during the exposure of the film.

To a very considerable extent the use of the kinematograph in schools, and the more general introduction of portable kinematographs [2] for home and commercial purposes, is dependent upon the development of a durable non-inflammable or slow-burning film. If slow-burning non-standard film is used, the choice of subjects is at the present time greatly restricted.

J. W. F.

[1] Carl Forch, *Der Kinematograph*, chap. xi. p. 135.

[2] A. F. Victor, "The Portable Projector," *Trans. Soc. Mot. Pict. Engrs.*, April 1918, p. 20.

Kirchhoff's Law of Radiation. See "Radiation Theory," § (4).

König-Martens Spectrophotometer. See "Spectrophotometry," § (12).

— L —

L Series: a group of spectrum lines in the characteristic X-rays emitted by an element. See "Radiology," § (17).

Lambert: a unit of brightness used in America. It is equal to the brightness of a perfectly diffusing surface of unity reflection ratio, illuminated to the extent of one phot (*q.v.*). It therefore equals $1/\pi$ candles per square centimetre. See "Photometry and Illumination," § (2).

Lamp Rotator: a device for obtaining a measurement of mean horizontal candle-power of a lamp by a single photometric reading. See "Photometry and Illumination," § (38).

Landolt and Lippich's Polarimeter: an instrument for measuring the position of the plane of polarisation, depending on the use of a very intense light source and a pair of wide-angled Nicol prisms, the field being not uniformly dark but crossed only by a narrow black band. See "Polarimetry," § (9).

Lantern Projection of Images on a Screen. See "Projection Apparatus," § (14).

Latitude, definition of. See "Surveying and Surveying Instruments," § (4).

Determination of, for survey purposes. See *ibid.* § (24).

Laurent Polarimeter. See "Polarimetry," § (5).

Lead, Use of, in Glass Manufacture. See "Glass," § (8).

Leading Marks for Fixing Position. See "Navigation and Navigational Instruments," § (2).

Length, Focal, determination of, for simple spherical lens. See "Lenses, The Testing of Simple."

Lens, aspherical: a type of spectacle lens designed to correct the aberration introduced when the eye looks towards the edge of the lens instead of to the centre. See "Ophthalmic Optical Apparatus," § (8) (ii.).

Bifocal: a type of spectacle lens having two different powers or foci set in the same eye-wire or spectacle frame; the upper focus is used for distance and the lower for reading or close work. See *ibid.* § (8) (i.).

Cataract: a type of spectacle lens for aphakic patients, designed to reduce the weight and aberration of the suitable lens of an ordinary type. See *ibid.* § (8) (iii.).

LENS SYSTEMS, ABERRATIONS OF

§ (1) Introduction.—The Gaussian theory of a system of coaxial lenses is given in the article on "Lenses, Simple Theory of," while in the articles on the Lenses for Telescopes,[3] Photographic Lenses,[4] and Microscopes [5] reference is made to the various aberrations to which any such system is subject.

In the following sections the theory is given of the five first-order aberrations, generally known by the name of Von Seidel, viz. Distortion, Curvature, Astigmatism, Coma, and Spherical Aberration. A more complete treatment will be found in the article on Systems of Lenses.

§ (2) Hamilton's Characteristic Function.—Light is propagated by transverse waves which spread outwards from any luminous source; in an isotropic medium—one, that is, which has identical properties in all directions—the wave velocity is the same in all directions, and the waves emanating from any point source spread outwards in sphere. Consider now the disturbance leaving the source at any given instant; at any future instant, supposing no reflection or refraction to have occurred, its effect will be spread over a sphere having the source as centre, and the

[3] See "Telescopes," § (3).

[4] See "Photographic Lenses," § (8).

[5] See "Microscope, The Optical Theory of," §§ (5), etc.

time taken to reach any point on this sphere will be the same, and will be given by dividing the radius of the sphere by the velocity of the light. This sphere constitutes the wave front for that disturbance. The effect of reflections or refractions, such as take place at the surfaces of a series of lenses, is to modify the form of the wave surface, but it always retains this fundamental property; it is the assemblage of points at which the disturbance emanating from the source at any given instant has arrived at the instant considered, and its form is given by the fact that the time taken to travel from the source to all points on the surface is the same. Moreover, it can be shown (i.) that the light which reaches any point is travelling at right angles to the wave surface through that point, and (ii.) that the ray or path by which the light has travelled is such that the time of transit from the source to the point is the least possible.

Now if x, y, z be the co-ordinates of any point on the wave front we may write its equation as $f(xyz)=c$, and since the time of transit from the origin to any point on the wave is constant we may express the time as a function of the variables which is constant when the point xyz is on the wave front. This will be satisfied if the time is expressed as a function of $f(xyz)$. Thus, if V denote the time we may put

$$V=\phi\{f(xyz)\}.$$

Again the direction cosines of the direction of propagation, which, as we have seen, is normal to the wave front, are proportional to df/dx, df/dy, and df/dz respectively. But

$$\frac{dV}{dx}=\frac{dV}{df}\frac{df}{dx}, \text{ etc.}$$

Thus the wave travels in a direction whose direction cosines, given say by L, M, N, are proportional to dV/dx, dV/dy, and dV/dz, when V represents the time taken to travel from the source—the origin—to the point x, y, z.

From this it follows that if V be regarded as a function of the co-ordinates of two points, one in the object, the other in the image, space—V being the time of transit between the points—the differential coefficients of V with regard to the co-ordinates of the points give the initial and final directions of the ray passing through the points. Thus a knowledge of V, which is the characteristic function of Hamilton, enables all the properties of the lens system to be deduced.

§ (3) Hamilton's Function for a Series of Coaxial Lenses.—The problem is simplified in the case of a system of lenses by the facts (i.) that the system is symmetrical about the axis of the lenses, and (ii.) that the objects and images considered are assumed to lie either in planes perpendicular to this axis or on surfaces of revolution about the axis.

Let X, Y, Z be the co-ordinates of a point P (*Fig.* 1) in the object space, consider a fixed

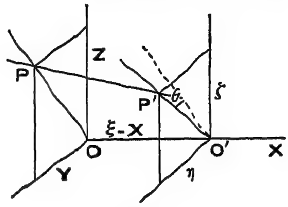


Fig. 1.

plane—the plane of a stop in the lens system—in the image space, and let the ray considered meet this in P′ whose co-ordinates are ξ, η, ζ.

Let planes through P and P′, normal to the axis, meet it in O, O′ respectively. Then it is clear from the symmetry about the axis that for any pair of points P, P′ in these planes V will be the same, so long as the distances OP, O′P′ and the angle between OP and O′P′ are fixed; thus, denoting this angle by θ, V may be treated as a function of the distance between the planes considered, the radii OP and O′P′ and the angle θ. But

$$OP=(Y^2+Z^2)^{\frac{1}{2}},$$

$$O'P'=(\eta^2+\zeta^2)^{\frac{1}{2}},$$

$$\cos\theta=\frac{Y\eta+Z\zeta}{OP.\,O'P'},$$

and $$OO'=\xi-X.$$

Thus we may treat V as a function of Y^2+Z^2, $\eta^2+\zeta^2$, $Y\eta+Z\zeta$, and $\xi-X$; this latter quantity is constant if the images are plane. Again in lens problems the quantities Y, Z, η, ζ are small compared with the focal lengths and other constants of the lens and with the distances of the object and image from the unit planes of the lens.

Thus we may expand V in powers of Y^2+Z^2, $Y\eta+Z\zeta$, and $\eta^2+\zeta^2$, and if we retain only the fourth powers of the small quantities we obtain the expression

$$\left.\begin{aligned}V=v_0&-\tfrac{1}{2}v_1(Y^2+Z^2)+v_2(Y\eta+Z\zeta)\\&-\tfrac{1}{2}v_3(\eta^2+\zeta^2)-\tfrac{1}{8}[v_4(Y^2+Z^2)^2\\&-4v_5(Y^2+Z^2)(Y\eta+Z\zeta)+4v_6(Y\eta+Z\zeta)^2\\&+2(v_6+v_7)(Y^2+Z^2)(\eta^2+\zeta^2)-4v_8(Y\eta+Z\zeta)(\eta^2+\zeta^2)\\&+v_9(\eta^2+\zeta^2)^2]+\ \ldots\ \text{terms involving higher}\\&\text{powers of the variables,}\end{aligned}\right\}\quad(1)$$

where the v's are quantities which depend only on the properties of the lenses through which the light has passed and on the positions of the object and stop planes, the values, that is, of X and ξ.

Also we have for the direction cosines of the emergent ray

$$M=\frac{dV}{d\eta},\quad N=\frac{dV}{d\zeta},\quad L=(1-M^2-N^2)^{\frac{1}{2}}.$$

Thus

$$\left.\begin{aligned}M=Y\{v_2+\tfrac{1}{2}v_5(Y^2+Z^2)-v_6(Y\eta+Z\zeta)\\+\tfrac{1}{2}v_8(\eta^2+\zeta^2)+\ .\ .\ .\ \}\\-\eta\{v_3+\tfrac{1}{2}(v_6+v_7)(Y^2+Z^2)-v_8(Y\eta+Z\zeta)\\+\tfrac{1}{2}v_9(\eta^2+\zeta^2)+\ .\ .\ .\ \}\end{aligned}\right\}, \quad (2)$$

with a similar expression for N, but with Z and ζ before the two brackets in place of Y and η. Also

$$L=1+\tfrac{1}{2}v_2^2(Y^2+Z^2)-v_1v_2(Y\eta+Z\zeta)+\tfrac{1}{2}v_3(\eta^2+\zeta^2)+\ .\ .\ . \quad (3)$$

§ (4) Position of the Image.—Now let the image plane be at a distance X′ from the stop plane, and let Y′, Z′ be the co-ordinates of the point in which the ray L, M, N meets the image plane. Then

$$\left.\begin{aligned}Y'&=\eta+\frac{M}{L}X'\\Z'&=\zeta+\frac{N}{L}X'\end{aligned}\right\}.\quad .\ .\ .\ (4)$$

Therefore

$$\left.\begin{aligned}Y'=\eta+MX'\{1-\tfrac{1}{2}v_2(Y^2+Z^2)\\+v_2v_3(Y\eta+Z\zeta)-\tfrac{1}{2}v_3(\eta^2+\zeta^2)\}\\=X'Y\{v_2+\tfrac{1}{2}(v_5-v_2^3)(Y^2+Z^2)\\-(v_6-v_2^2v_3)(Y\eta+Z\zeta)\\+\tfrac{1}{2}(v_8-v_2v_3^2)(\eta^2+\zeta)^2\}\\+X'\eta\{(1/X')-v_3\\-\tfrac{1}{2}(v_6-v_2^2v_3+v_7)(Y^2+Z^2)\\+(v_8-v_2v_3^2)(Y\eta+Z\zeta)\\-\tfrac{1}{2}(v_9-v_3^3)(\eta^2+\zeta^2)\}\end{aligned}\right\},\quad (5)$$

with a similar equation for Z′.

Now the condition that the plane for which the intersections have been found is the correct image plane, is that, for terms of the lowest order, Y′ and Z′ shall be proportional to Y and Z respectively, for the image must be similar to the object. Thus we must have $X'=1/v_3$. Hence the traces of the ray intersections with the image plane are found by substituting this value of X′ in the above expressions.

We can simplify these expressions by supposing that the axis of y passes through the source of the light so that $Z=0$. If this be done, we find the values

$$\left.\begin{aligned}2v_3Y'=2v_2Y+(v_5-v_2^3)Y^3\\-\{3(v_6-v_2^2v_3)+v_7\}Y^2\eta\\+(v_8-v_2v_3^2)Y(3\eta^2+\zeta^2)\\-(v_9-v_3^3)\eta(\eta^2+\zeta^2)\\2v_3Z'=-\{(v_6-v_2^2v_3)+v_7\}Y^2\zeta\\+2(v_8-v_2v_3^2)Y\eta\zeta\\-(v_9-v_3^2)\zeta(\eta^2+\zeta^2)\end{aligned}\right\}.\quad (6)$$

§ (5) The Aberrations.—These expressions agree with those employed in discussing the aberrations separately in § (8) of the article on Photographic Lenses if we put

$$v_2Y=v_3y,$$
$$Y'=y+\delta y,$$
$$Z'=0+\delta z,$$

so that y, 0 are the co-ordinates of the image when there is no aberration, and δy, δz the aberrations in the radial and transverse directions respectively.

(i.) *Distortion.*—The first term in the expression for the aberrations is given by

$$\delta y=\frac{v_5-v_2^3}{2v_3}Y^3,\quad .\ .\ .\ .\ (7)$$

and this gives the distortion [1] in the form

$$\delta y=a_1y^3,$$

where $$a_1=\frac{(v_5-v_2^3)}{2v_2^3}v_3^2\,;$$

the image is displaced by this amount in the radial direction and the aberration is wholly radial.

(ii.) *Curvature and Astigmatism.*—We have next the terms in $Y^2\eta$ and $Y^2\zeta$ given by

$$\left.\begin{aligned}2\delta y&=-\frac{1}{v_3}\{3(v_6-v_2^2v_3)+v_7\}Y^2\eta\\ \text{and}\quad 2\delta z&=-\frac{1}{v_3}\{(v_6-v_2^2v_3)+v_7\}Y^2\zeta\end{aligned}\right\},\ .\ (8)$$

or, as we may write them,[2] remembering that

$$Y^2=\frac{v_3^2}{v_2^2}y^2,$$

$$\frac{2\delta y}{y^2}=\eta(3e+e'),$$

and $$\frac{2\delta z}{y^2}=\zeta(e+e'),$$

where $$e=\frac{v_3}{v_2^2}(v_6-v_2^2v_3),$$

$$e'=\frac{v_3}{v_2^2}v_7.$$

To examine the nature of this aberration more closely consider the intersection of the rays with a sphere given by the equation

$$X'=\frac{1}{v_3}+\tfrac{1}{2}\left(\frac{Y'^2+Z'^2}{R}\right),$$

where R is the radius of the sphere.

Additional terms are thereby introduced into the right of equation (5), which are given by

$$\frac{1}{2R}\left(\frac{v_2}{v_3}\right)^2(Yv_2-\eta v_3)(Y^2+Z^2),$$

or, putting $Z=0$, the further terms in $2v_3Y'$ are

$$\frac{1}{R}\frac{v_2^3}{v_3}Y^3-\frac{1}{R}v_2^2Y^2\eta,$$

and in $2v_3Z'$

$$-\frac{1}{R}v_2^2Y^2\zeta.$$

[1] See "Photographic Lenses," § (8) (ii.) (*a*).
[2] See *ibid.* § (8) (ii.) (*c*).

The additional term in Y^3 means that the ray through the centre of the stop for which $\eta=\zeta=0$ strikes the sphere in a point at a different distance from the axis from that in which it meets the plane.

The complete coefficients of the terms in $Y^2\eta$ and $Y^2\zeta$ become

$$-\left\{3(v_6-v_2{}^2v_3)+v_7+\frac{1}{R}v_2{}^2\right\}$$

and $$-\left\{(v_6-v_2{}^2v_3)+v_7+\frac{1}{R}v_2{}^2\right\}.$$

Thus the spreading of the light which these terms imply disappears if surfaces of curvature

$$-\frac{1}{v_2{}^2}\{3(v_6-v_2{}^2v_3)+v_7\}$$

and $$-\frac{1}{v_2{}^2}\{(v_6-v_2{}^2v_3)+v_7\}$$

be considered as the image surfaces instead of the plane given by $X'=1/v_3$. In other words, the defects are due to the fact that pencils for which η is appreciable and ζ small come to a focus on a spherical surface having the former curvature, those for which η is small and ζ appreciable focus on a surface having the latter curvature. Thus the images are curved. Clearly, also, if e is finite the two spheres are distinct, no point image is formed; the aberration to which this gives rise is known as astigmatism. These curvatures may be written $-(3e+e')/v_3$ and $(e+e')/v_3$, and should be compared with the corresponding expressions given in the articles on "Optical Calculations," § (17), and "Eyepieces," § (4).

(iii.) *Coma.*—The next two terms[1] in the expressions for δy and δz are

$$\left.\begin{aligned}2v_3\delta y&=(v_8-v_2v_3{}^2)(3\eta^2+\zeta^2)Y\\ 2v_3\delta z&=2(v_8-v_2v_3{}^2)\eta\zeta Y\end{aligned}\right\}, \quad (9)$$

and these may be written

$$\frac{\delta y}{cy}=2(\eta^2+\zeta^2)+(\eta^2-\zeta^2)$$

$$=r^2(2+\cos 2\theta)$$

and $$\frac{\delta z}{cy}=2\eta\zeta=r^2\sin 2\theta;$$

if $$\eta=r\cos\theta,\ \zeta=r\sin\theta,$$

$$\therefore(\delta y-2cyr^2)^2+(\delta z)^2=c^2y^2r^4.$$

Thus the rays from a zone of the stop for which r^2 is equal to a constant pass through a circle in the image plane, the co-ordinates of whose centre are given by $2cyr^2$, 0, and whose radius is cyr^2.

It will be noticed that $\delta z=0$, *i.e.* the rays meet the image plane in the axial plane $z=0$, both when $\eta=0$ and when $\zeta=0$. In the first case

$$\frac{\delta y}{cy}=\zeta^2=r^2$$

and in the second

$$\frac{\delta y}{cy}=3\eta^2=3r^2.$$

Thus, in the second case the displacement of the image from its ideal position is three times as great as in the first.

Again the rays from opposite ends of any diameter of the annulus of the stop, *i.e.* from points whose co-ordinates are respectively η, ζ, and $-\eta$, $-\zeta$, meet the image plane in the same point. The complete circle in the image plane is formed by rays from half the annulus of the stop, thus when the complete annulus is considered the image circle is formed twice over.

(iv.) *Spherical Aberration.*—Spherical aberration[2] is independent of Y; it exists when Y is zero, or the light comes from a point on the axis of the system. In the expressions for $2v_3Y'$ and $2v_3Z'$ we have the terms

$$-(v_9-v_3{}^3)\eta(\eta^2+\zeta^2)$$

and $$-(v_9-v_3{}^3)\zeta(\eta^2+\zeta^2),$$

or, as we may write them,

$$\frac{\delta y}{\eta}=\frac{dz}{\zeta}=v_1r^2. \quad . \quad . \quad . \quad (10)$$

These do not vanish with Y, and indicate that a ray from an axial point passing through a point $\eta\zeta$ of the stop is not brought to a focus at the same point as a central ray for which X' is $1/v_3$, but at a point for which X' has the value given by

$$\frac{1}{v_3}+\tfrac{1}{2}S(\eta^2+\zeta^2),$$

where S is a quantity depending on the spherical aberration.

We find on making this assumption an additional term in the expression for $2v_3Y'$,

$$\tfrac{1}{2}S(v_2Y-v_3\eta)(\eta^2+\zeta^2),$$

and in that for $2v_3Z'$,

$$-\tfrac{1}{2}Sv_3\zeta(\eta^2+\zeta^2).$$

The terms in $(\eta^2+\zeta^2)$ then disappear from our expressions for δy and δz if

$$\tfrac{1}{2}Sv_3+(v_9-v_3{}^3)=0,$$

i.e. if $$S=\frac{2(v_9-v_3{}^3)}{v_3},$$

which indicates different focussing planes for light from an axial source passing through

[1] See "Photographic Lenses," § (8) (ii.) (*c*).

[2] See "Photographic Lenses," § (8) (ii.) (*b*).

distinct zones of the slot. The axial displacement δx of the focus is given by

$$\delta x = \mathrm{R}r^2,$$

where, as before, r is the radius of the zone.

In applying this analysis of the aberration to any actual system it must be borne in mind that, in order to obtain a satisfactory correlation between theory and the appearance of the image, the intensity distribution derivable from theory must be considered in conjunction with the geometrical trace, and, further, that the theory must be modified to allow for the effects of diffraction whenever very small dimensions are in question. The appearances due to aberration are modified from those derivable from the foregoing theory of small aberrations.

LENSES, concave, the testing of. See "Lenses, The Testing of Simple," § (3).
Convex, the testing of. See *ibid.* § (1).
Ophthalmic. See "Ophthalmic Optical Apparatus," § (8).
Testing of Camera. See "Camera Lenses, Testing of."
Thin, the theory of, and application of results to "close' lenses. See "Optical Calculations," § (7).

LENSES, THE TESTING OF SIMPLE.

THE chief characteristics to be examined in the testing of a simple spherical lens are, firstly, the focal length, and secondly, the centering error, that is to say, any want of coincidence of the optical centre with the geometrical centre of the lens. In the case of a cylindrical lens attention must also be paid to the orientation of the axis of the cylinder.

The theory on which the tests are based is dealt with elsewhere.[1]

These tests are most easily carried out on an optical bench such as is illustrated in *Figs.* 1, 2, and 3. A standard at one end of the

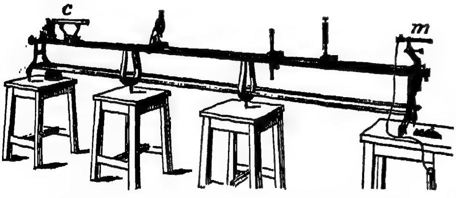


FIG. 1.—Optical Bench for testing the Lenses in Oculists' Trial Cases.

bench holds a collimator c (*Fig.* 1) with illuminated vertical and horizontal cross-wires at its focus. At the other end, facing towards the collimator, is a microscope m which can be moved parallel to itself in a vertical or horizontal direction, while at its focus is a graticule M (*Fig.* 2) engraved with vertical

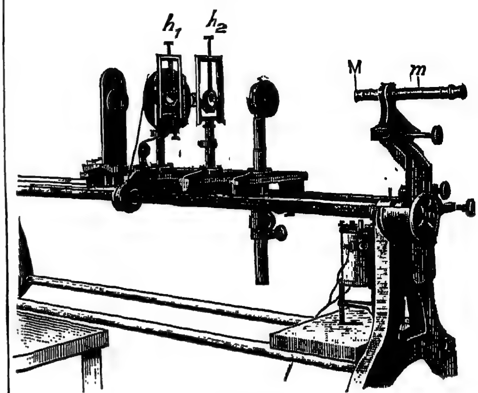


FIG. 2.

and horizontal cross-lines cutting symmetrically across a series of concentric circles of radii 1, 2, 3 . . . mm. An endless steel tape t (*Fig.* 3), attached to a lens-holder h_1, and passing over a pulley at each end of the bench, enables the observer to adjust the holder to

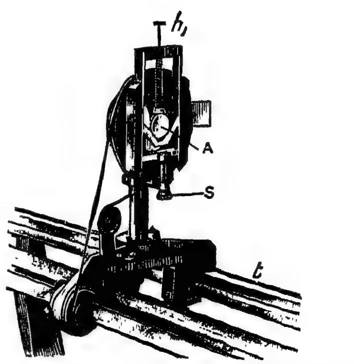


FIG. 3.—Travelling Carriage with Rotating Head on Trial Case Lens Testing Bench.

any position along the bench. The portion of the holder which grips the lens can be rotated about a horizontal axis in line with the axis of the collimator. The lens A, which is held in position by an adjustable screw s, can thus be rotated in its own plane, the amount of rotation being indicated on a circular scale. A second standard h_2 for holding an auxiliary lens can be placed on the bench at any point and its position read by means of a pointer working over a scale on the bench.

(i.) *Convex Lenses.*—The lens under test is placed in the holder h_1 and brought close

[1] See "Lenses, Theory of Simple."

up to the microscope, the position of which is adjusted to bring one edge of the lens into the field of view. The screw holding the lens is adjusted so that on rotation the edge of the lens remains in a fixed position relative to the microscope cross-lines; the geometrical centre of the lens is then on the axis of the bench. Next, the holder is racked along the bench until the lens focusses the collimator cross-wires on the graticule of the microscope. If the lens is rotated, any separation of the optical centre from the geometrical centre will cause a corresponding rotation of the central point of the image in a circle of radius equal to the centering error in the lens, and by adjusting the microscope to make this circle of rotation concentric with the circles engraved on the graticule, whose radii are known, the centering error can be estimated to 0·1 mm. If, as in cases considered later, an auxiliary lens is used between the test lens and the microscope, the magnification it produces must be taken into account when estimating the centering error by this method.

When the lens A focusses the parallel rays from the collimator on to the graticule M (*Fig.* 4), the distance AM represents the focal length of the lens, neglecting the thickness of the latter. The position of A can be read on the tape, and the corresponding reading for M may be found by taking the tape reading when the lens is separated from the graticule by a small gauge of known length; the difference between these readings for A and M will give the distance of the focus from the back surface of the lens, that is, the back focal length. If the lens is thin, this will not differ from the true focal length by more than the probable experimental error. All plano-convex or plano-concave lenses should be tested with their curved surface towards the microscope, for then the second nodal point coincides with this surface and the focal length will be equivalent to the back focal length.

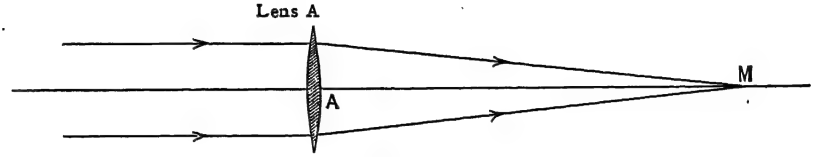


Fig. 4.

For biconvex lenses of equal radii and of considerable thickness a length[1] $2t/\mu$ must be added to the back focal length to give the true focal length, t being the central thickness and μ the refractive index of the lens. In obtaining this expression the square of the thickness has been omitted. For glass of refractive index 3/2 the correction comes to $t/3$.

(ii.) *Use of Auxiliary Lens.*—This method of measuring focal lengths is obviously applicable only to positive lenses whose focal length does not exceed the available length of the bench; with lenses of longer focus and with negative lenses a positive auxiliary lens must be used. The arrangement is indicated in *Figs.* 5 and 6. An equiconvex lens B of suitable focal length is mounted on the bench and its position noted when it brings parallel rays to a focus at the graticule M (*Fig.* 5). If F_1 and F_2 represent the foci of lens B, F_2 will in this case coincide with M. Let the distance B_2F_2, determined as described before, be p. The test lens is then introduced between lens B and the collimator (*Figs.* 6A, 6B), and the position of each lens is read when together they bring the rays to a focus at M. Let the distances AM and B_2M be now q and r respectively. The formation of the image is shown for a positive lens in *Fig.* 6A and for a negative lens in *Fig.* 6B. In *Fig.* 6A parallel rays passing through A converge towards O; O may therefore be taken as a virtual object the image of which is produced by the lens B at M. If B is equiconvex and of thickness t, assuming a refractive index of 3/2, its first and second nodal points lie within the lens at a distance[1] $t/3$ from B_1 and B_2 respectively; hence the focal length of B is $p+t/3$, while the distances of the object O from the first nodal point, and of the image M from the second nodal point, are respectively $B_2O+2t/3$ and $B_2M+t/3$.

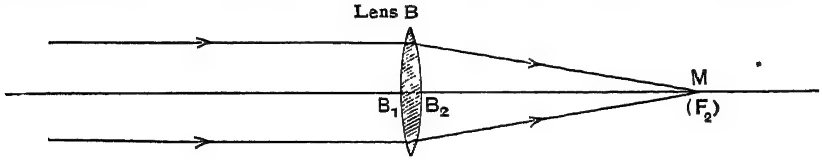


Fig. 5.

[1] See "Lenses, Theory of Simple."

Therefore, using the ordinary formula $f(u-v)=uv$, we find

$$\left(p+\frac{t}{3}\right)\left(B_2O+\frac{2t}{3}-B_2M-\frac{t}{3}\right)=\left(B_2O+\frac{2t}{3}\right)\left(B_2M+\frac{t}{3}\right)$$

or

$$\left(p+\frac{t}{3}\right)\left(f-q+\frac{t}{3}\right)=\left(f-q+\frac{t}{3}+r+\frac{t}{3}\right)\left(r+\frac{t}{3}\right),$$

giving

$$\left(f-q+\frac{t}{3}\right)(p-r)=\left(r+\frac{t}{3}\right)^2.$$

focal length of the auxiliary lens and depends only upon the amount of shift of this lens between the two positions, a measurement much less susceptible of error. The second formula is also independent of the thickness of the auxiliary lens; if necessary, the formula may be modified to take account of the thickness of lens A.

(iii.) *Concave Lenses.*—A third method is applicable for negative lenses only; the arrangement adopted is indicated in *Fig.* 8. The lens B is placed in any convenient position so that lens A may first be adjusted to produce in conjunction with B an image of the colli-

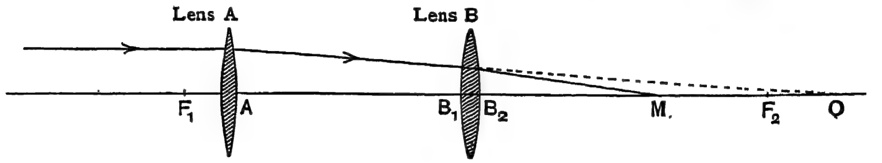


FIG. 6A.

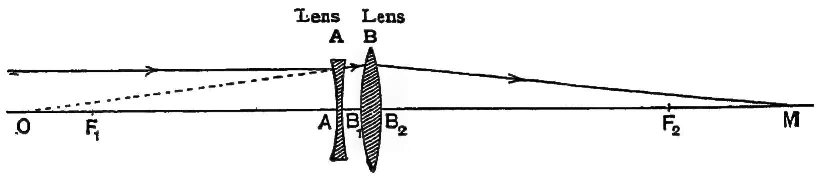


FIG 6B.

Hence the focal length of lens A

$$=f=q-\frac{t}{3}+\frac{\left(r+\frac{t}{3}\right)^2}{p-r}.$$

A similar formula is obtained in the case of a negative lens, illustrated by *Fig.* 6B.

Another arrangement is sometimes preferable. In this lens A is placed between the

mator cross-wires at M, and afterwards may be placed at the conjugate point to M with respect to the lens B; in *Fig.* 8, A_1 and A_2 represent these two positions. The position A_2 is determined by dusting a little lycopodium powder on to the back surface of A, and adjusting A until an image of the powder is focussed by lens B at M. The distance moved by lens A between the two positions is equivalent

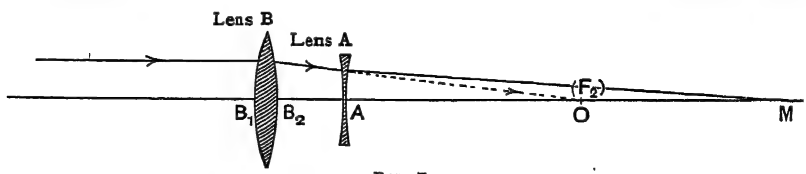


FIG. 7.

auxiliary lens and the microscope (*Fig.* 7). Using the same notation as before, the same formula may be applied, this time connecting the distances of O and M from the lens A. Neglecting the thickness of A, we have

$$f(AM-AO)=AM \cdot AO$$

or
$$f=\frac{q(p+q-r)}{r-p}=\frac{q^2}{r-p}-q.$$

It will be noted that with both arrangements the focal length f is independent of the actual

to the back focal length; the correction to be applied to obtain the true focal length is very small unless the central thickness of the lens is large.

The accuracy obtainable by these methods is not very great, in the case of long focus lenses especially, but the limits of error are small compared with those which are permissible in a simple lens used for a spectacle lens, and it is for lenses of this class that the tests are generally used. More accurate methods of focal length measurement, which

may also be applied to simple lenses, will be found under the article "Objectives, The Testing of Compound."

(iv.) *Optical Standards Committee*.—In 1908 the Optical Standards Committee of the Optical Society laid down certain limits for the error permissible in the manufacture of spectacle lenses;[1] these have obtained recognition generally, and are the limits allowed in the testing of trial case lenses at the National Physical Laboratory.

As the thickness of a negative lens may be kept small, no matter what the power may be, the latter can be found with considerable precision by measurement of the surface curvatures, taking a standard value of 1·52 for the refractive index of the glass. In the case of positive lenses the power is less independent of the thickness, especially with strong lenses; it is the recognised custom, therefore, to make the standard power and thickness of a positive lens such that it will exactly neutralise a standard negative lens of equal power. This method of neutralisation gives a simple means of checking one lens against another, and with experience an accuracy of at least $\frac{1}{8}$ D may be obtained.

According to the rules adopted by the Committee, the true power of a negative lens must not differ from the nominal value marked on it by more than ±2 per cent, or by more than ±0·01 D if this is greater than 2 per cent. For positive lenses of powers less than 10 D the limits are the same; for equiconvex lenses between 10 D and 12 D inclusive the limits of error are +3 per cent and −1 per cent, and for lenses between 13 D and 20 D inclusive the limits are +2½ per cent and −1 per cent.

The standard thickness of all negative lenses is 1 mm. at the centre. The standard thickness of a positive lens of any definite power is the minimum thickness possible for a lens of that power with a diameter of 3·8 cm. (1½ in.) and an edge thickness of 0·5 mm., the refractive index being 1·50.[2] This varies from 4·3 mm. for a 10 D lens to 7·8 mm. D for a 20 D lens.

The limits of error allowable in the centering of a spherical lens are as follows:

1 mm. for lenses between 20 D and 1 D.
2 mm. for lenses between 0·87 D and 0·37 D.
3 mm. for a lens of 0·25 D.
5 mm. for a lens of 0·12 D.

The maximum permissible error in the marking of the cylindrical axis of a lens is 2°; on a 3·8 cm. (1½ in.) lens this is equivalent to a displacement of each of the marks by 0·6 mm.

A. B. D.

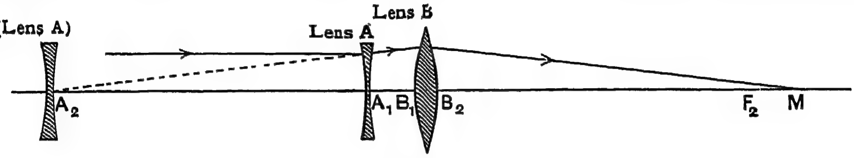


FIG. 8.

[1] Report of Optical Standards Committee on Standardisation of Trial Cases, 1908.

[2] An increase of 1 mm. in the thickness of a lens produces a change of power of −0·06 D in a lens of power 20 D, and −0·02 D in a lens of 10 D.

LENSES, THEORY OF SIMPLE

§ (1) GAUSS THEORY OF LENSES.—The simple theory of lenses developed below is due to Gauss. It is assumed that the rays of light considered are nowhere inclined at more than a small angle to the axis of the lens; that to any point in the object space there corresponds a point-image in the image space, and that the image of a small line in the object space intersecting the axis at right angles is itself a small line at right angles to the axis in the image space.

We assume further, in obtaining the formulae, that light is travelling from left to right and that lines drawn in this direction are positive; the radius of any spherical surface on which the light falls is thus positive when the surface is convex to the light.

§ (2) LENS FORMULAE.—Consider a ray PQ (*Fig.* 1) converging to a point Q on the axis AO

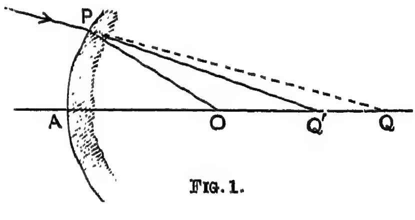


FIG. 1.

of spherical surface AP; it is refracted at the surface and the refracted ray cuts the axis at Q′, and if $AQ=u$, $AQ'=v'$, $AO=r$, and the refractive index be μ, we have

$$\frac{\mu}{v'}-\frac{1}{u}=\frac{\mu-1}{r}.$$

For $$\mu=\frac{\sin OPQ}{\sin OPQ'}=\frac{\sin OPQ}{\sin POQ}\cdot\frac{\sin POQ'}{\sin OPQ'}$$

$$=\frac{OQ}{PQ}\times\frac{PQ'}{OQ'}=\frac{OQ}{AQ}\times\frac{AQ'}{OQ'},$$

when P is very near to A.

Hence $\mu u(v'-r)=v'(u-r)$

and $$\frac{\mu}{v'}-\frac{1}{u}=\frac{\mu-1}{r}. \quad . \quad . \quad . \quad (1)$$

Consider now the case of light traversing a biconvex lens ACB′ (*Fig.* 2); the ray PQ′ strikes

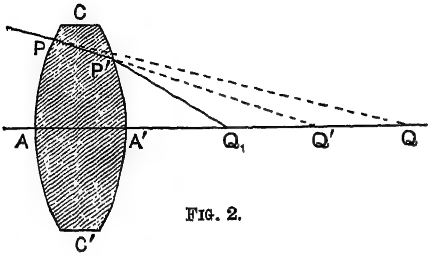


FIG. 2.

the concave surface at P′ and is again refracted, cutting the axis at Q_1. Let $A'Q_1=v_1$, the radius of the second surface be s and the thickness AA′ be t. We assume the medium on both sides of the lens to be the same.[1] Since the surface is concave to the light s is negative. Also $A'Q'=v'-t$, and we have, the refractive index from glass to air being $1/\mu$,

$$\frac{1/\mu}{v}-\frac{1}{v'-t}=-\frac{(1/\mu)-1}{s}=\frac{\mu-1}{\mu s},$$

or $$\frac{1}{v}-\frac{\mu}{v'-t}=\frac{\mu-1}{s}. \quad . \quad . \quad (2)$$

On substituting in (2) the value of v' derived from (1) and reducing we arrive at an equation which may be written either as

$$v=\frac{\gamma u+\delta}{\alpha u+\beta}, \quad . \quad . \quad . \quad (3)$$

or $$u=\frac{\delta-\beta v}{\alpha v-\gamma}, \quad . \quad . \quad . \quad (4)$$

where

$$\left.\begin{aligned}\alpha&=(\mu-1)\left(\frac{1}{r}+\frac{1}{s}\right)-\frac{(\mu-1)^2}{rs}\frac{t}{\mu}\\&=(\mu-1)\left(\frac{1}{r}+\frac{1}{s}\right)\left\{1-\frac{\mu-1}{r+s}\frac{t}{\mu}\right\},\\\beta&=1-\frac{\mu-1}{s}\frac{t}{\mu},\\\gamma&=1-\frac{\mu-1}{r}\frac{t}{\mu},\\\delta&=-\frac{t}{\mu}.\end{aligned}\right\} \quad . \quad (5)$$

By substituting the values of α, β, γ, δ, found above, we can also prove that

$$\alpha\delta-\beta\gamma=-1. \quad . \quad . \quad . \quad (6)$$

§ (3) PRINCIPAL FOCI.—If in the above formulae we make $u=\infty$ so that the incident pencil consists of parallel rays, we find

$$v=\frac{\gamma}{\alpha}=A'F' \text{ (say).}$$

F′ (*Fig.* 5) is then the point known as the second principal focus.

While if we make $v=\infty$ so that the emergent pencil is parallel, then

$$u=-\frac{\beta}{\alpha}=AF \text{ (say),}$$

and F is the first principal focus.

If Q and Q′ are any pair of conjugate points, then

$$\begin{aligned}FQ.F'Q'&=(AQ-AF)(A'Q'-A'F')\\&=\left(u+\frac{\beta}{\alpha}\right)\left(v-\frac{\gamma}{\alpha}\right),\end{aligned}$$

and this, on substituting for u and v, leads to

$$FQ.F'Q'=\frac{\alpha\delta-\beta\gamma}{\alpha^2}=-\frac{1}{\alpha^2}. \quad . \quad . \quad (7)$$

§ (4) MAGNIFICATION CAUSED BY REFRACTION.—We can find an expression for the magnification caused by refraction thus. Consider the rays converging to a small object QM (*Fig.* 3) at right angles to the axis of a single refracting surface AP with its centre at

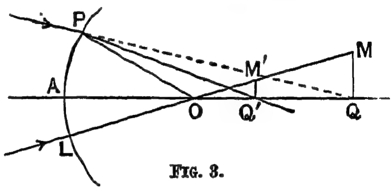


FIG. 3.

O. An image M′Q′ is formed at Q′ and with the previous notation

$$\frac{\mu}{v'}-\frac{1}{\mu}=\frac{\mu-1}{r}.$$

An incident ray such as LOM, which passes through the centre O and also through a point M of the object, passes on without deviation. Thus M′, the image of M, lies on this ray; it will therefore be the point where a line Q′M′ drawn through Q′ at right angles to the axis cuts LOM. Thus Q′M′ is the image of QM, and the magnification m_1 due to the one refraction is measured by the ratio of Q′M′ to QM. Hence

$$m_1=\frac{Q'M'}{QM}=\frac{v'-r}{u-r}.$$

But we have already seen, § (2), that $\mu u(v'-r)=v'(u-r)$. Thus

$$m_1=\frac{1}{\mu}\frac{v'}{u}. \quad . \quad . \quad . \quad (8)$$

Proceeding now to the case of the lens (*Fig.* 4), the image Q′M′ formed by the first refraction is again magnified[2] at the second

[1] For the more complicated case of a series of lenses see "Optical Calculations."

[2] In the figure as drawn the image is diminished, not magnified; the magnification is less than unity.

surface, and clearly we have similarly for this magnification m_2 the value $m_2=\mu v/(v'-t)$.

The resultant magnification m is the product of the two quantities m_1 and m_2. For

$$m=\frac{Q_1M_1}{QM}=\frac{Q'M'}{QM}\times\frac{Q_1M_1}{Q'M'}=m_1\times m_2.$$

Thus $$m=\frac{v'}{u}\times\frac{v}{v'-t}=\frac{v}{u}\cdot\frac{v'}{v'-t}.$$

On substituting in this the value of v' found

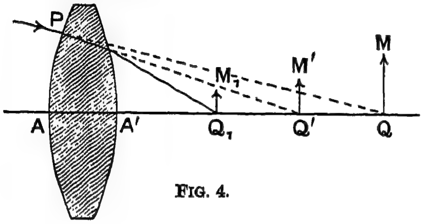


FIG. 4.

from equation (1) and using (3) and (4), we obtain

$$m=\frac{1}{\beta+\alpha u}=\gamma-\alpha v. \quad . \quad . \quad (9)$$

§ (5) PRINCIPAL POINTS OR POINTS OF UNIT MAGNIFICATION.—If, in the expressions for the magnification, we put m equal to unity so that the magnification is unity, the image and object are of the same size, and we find

$$\beta+\alpha u=1,$$

or $$u=\frac{1-\beta}{\alpha}, \quad . \quad . \quad . \quad (10)$$

and $$\gamma-\alpha v=1,$$

or $$v=\frac{\gamma-1}{\alpha}. \quad . \quad . \quad . \quad (11)$$

The two points thus defined are points of unit magnification; they are conjugate points

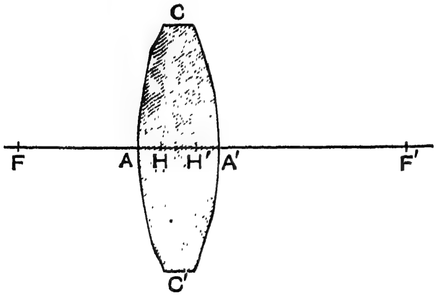


FIG. 5.

on the axis; we denote them by H and H′ respectively (*Fig.* 5). They are known as the unit or principal points of the lens, and an image of a small object at right angles to the axis at one principal point will be formed at the other, and will be equal in size to the object.

The first principal focus is, we have seen, given by the equation

$$AF=-\frac{\beta}{\alpha},$$

and the first principal point H by

$$AH=\frac{1-\beta}{\alpha}.$$

Thus $$HF=AF-AH=-\frac{1}{\alpha}, \quad . \quad . \quad (12)$$

which gives the distance between the first principal focus and the first unit or principal point.

Again, $$A'F'=\frac{\gamma}{\alpha},$$

and $$A'H'=\frac{\gamma-1}{\alpha}.$$

Thus $$H'F'=A'F'-A'H'=\frac{1}{\alpha}, \quad . \quad . \quad (13)$$

or the distances between the principal foci and the corresponding principal points are the same.

§ (6) LENS FORMULA WHEN THE DISTANCES ARE MEASURED FROM THE PRINCIPAL POINTS.—Let u_1 v_1 be the distances of the object and its image from the first and second principal points respectively, then

$$u=u_1+\frac{1-\beta}{\alpha},$$

$$v=v_1+\frac{\gamma-1}{\alpha}.$$

On substituting these values in the equation

$$v=\frac{\gamma u+\delta}{\alpha u+\beta},$$

we obtain

$$\frac{1}{v_1}-\frac{1}{u_1}=\alpha. \quad . \quad . \quad . \quad (14)$$

If we put $v=\infty$ so that the emergent pencil is parallel, we have

$$u_1=-\frac{1}{\alpha};$$

while if $u=\infty$ so that the incident pencil is parallel, we obtain

$$v_1=\frac{1}{\alpha}.$$

Thus as before

$$HF=-\frac{1}{\alpha}$$

and $$H'F'=\frac{1}{\alpha}.$$

The distance H′F′ or $1/\alpha$ measures the focal length of the lens.

§ (7) Nodal Points.—We can obtain expressions for the magnification in terms of the distances u_1 and v_1 from the principal points thus,

$$m=\frac{1}{\beta+\alpha u}=\frac{1}{\beta+\alpha u_1+(1-\beta)}$$

$$=\frac{1}{1+\alpha u_1}, \quad . \quad . \quad . \quad (15)$$

Or in terms of v_1

$$m=\gamma-\alpha\left(v_1+\frac{\gamma-1}{\alpha}\right)=1-\alpha v_1. \quad . \quad (16)$$

But $$\alpha=\frac{1}{v_1}-\frac{1}{u_1}.$$

Thus $$1-\alpha v_1=-\frac{v_1}{u_1}.$$

Hence $$m=-\frac{v_1}{u_1}.$$

Thus in *Fig.* 6

$$\frac{M'Q'}{MQ}=m=\frac{H'Q'}{HQ},$$

and the triangles MHQ and M'H'Q' are similar, so that MH is parallel to H'M' or a ray incident through the first principal

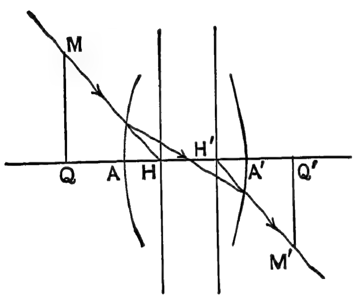


Fig. 6.

point emerges, parallel to its original direction, through the second. We have thus another property of the unit or principal points which are in consequence known as nodal points.

The coincidence of the principal and nodal points is a consequence of the assumption that the first and third media are the same. In general, points having the property of nodal points can be found, but they do not coincide with the principal points.

§ (8) Applications to Biconvex Lenses.—The values of the constants α, β, γ, δ in terms of the form and refractive index of the lens are given in equations (5).

Making use of these equations we obtain the following results for the positions of the principal foci and unit points neglecting powers of t above the first. We have

$$AF=-\frac{\beta}{\alpha}=-\frac{rs}{(\mu-1)(r+s)}\left\{1-\frac{\mu-1}{\mu}\cdot\frac{t}{r+s}\cdot\frac{r}{s}\right\}, \quad (17)$$

$$A'F'=\frac{\gamma}{\alpha}=\frac{rs}{(\mu-1)(r+s)}\left\{1-\frac{\mu-1}{\mu}\cdot\frac{t}{r+s}\cdot\frac{s}{r}\right\}, \quad (18)$$

$$AH=\frac{1-\beta}{\alpha}=\frac{r}{r+s}\cdot\frac{t}{\mu}, \quad . \quad . \quad (19)$$

$$A'H'=\frac{\gamma-1}{\alpha}=-\frac{s}{r+s}\,\frac{t}{\mu}, \quad . \quad . \quad (20)$$

while for the focal length $H'F'=f$, we have

$$f=H'F'=\frac{1}{\alpha}$$

$$=\frac{1}{\mu-1}\cdot\frac{rs}{r+s}\left\{1+\frac{\mu-1}{\mu}\cdot\frac{t}{r+s}\right\}. \quad (21)$$

If we put $t=0$ in these expressions we recover the ordinary formulae for a thin lens.

We may write the expression for the focal length as

$$f=\frac{1}{\mu-1}\cdot\frac{rs}{r+s}+\frac{rs}{(r+s)^2}\,\frac{t}{\mu}$$

$$=f_1+\frac{rs}{(r+s)^2}\,\frac{t}{\mu}, \quad . \quad . \quad . \quad . \quad (22)$$

where f_1 is the focal length of the lens treated as thin. In the case of an equiconvex lens of glass for which $r=s$, $\mu=3/2$ the value of the correcting term is $t/6$, while in the same case the principal points are within the lens and at a distance of $t/3$ from the vertices. If one surface, say the second, be flat, then s is infinite, the correction to the focal length is zero. The value of AH is also zero, so the first focal point is on the curved surface of the lens, while for A'H' we have the value $-t/\mu$ or, taking as before $\mu=3/2$, $A'H'=-2t/3$, the secondary principal point is within the lens at a distance of 2/3 of the thickness from the flat side.

§ (9) Biconcave Lenses.—The above formulae have been developed for the case of a biconvex lens in air. Any other lens can be treated similarly, having due regard to the signs of r and s.

Thus for a biconcave lens r is negative and s positive; we have therefore to change the signs of both r and s in the above formulae, and find

$$AF=\frac{rs}{(\mu-1)(r+s)}\left\{1+\frac{\mu-1}{\mu}\cdot\frac{t}{r+s}\cdot\frac{r}{s}\right\}, \quad (23)$$

$$A'F'=-\frac{rs}{(\mu-1)(r+s)}\left\{1+\frac{\mu-1}{\mu}\cdot\frac{t}{r+s}\cdot\frac{s}{r}\right\}, \quad (24)$$

$$AH=\frac{r}{r+s}\,\frac{t}{\mu}, \quad . \quad . \quad . \quad . \quad (25)$$

$$A'H'=-\frac{s}{r+s}\,\frac{t}{\mu}, \quad . \quad . \quad . \quad (26)$$

$$f=-\frac{1}{\mu-1}\,\frac{rs}{r+s}\left\{1-\frac{\mu-1}{\mu}\,\frac{t}{r+s}\right\}. \quad (27)$$

§ (10) General Cases.—Cases in which more than one lens occurs are dealt with by algebraical methods in the article on "Optical Calculations," to which the reader is referred. The practical application of the formulae to find the focal length and other properties of lenses is treated of in the articles "Lenses, Testing of Simple," and "Objectives, Testing of Compound," to which reference should be made. Microscope lenses are dealt with in the article "Microscope, Optics of the."

§ (11) Aberrations.—No lens behaves even for rays which diverge from a point on its axis in exactly the manner described, and when the object point is not on the axis the divergences from the above theory are more marked still.

The position[1] of the image point corresponding to a given object point can be expressed in terms of the distance of the object from the lens, of a series of ascending powers of the angle between the axis of the lens and the central ray of the pencil which forms the image, and of the curvatures, thickness, and refractive index of the lens. In the above discussion terms involving the third and higher powers of the angle of incidence have been neglected, symmetry ensures that no terms involving even powers of the angle occur. If, however, the higher powers be included the theory developed needs considerable corrections.[2] These faults in the geometrical theory of strict linear correspondence between the object and image space are known as aberrations.

If we limit ourselves to third powers of the obliquity of the central ray, the aberrations, as von Seidel[3] showed, are five in number, viz.:

(i.) *Spherical Aberration.*—Rays diverging from an object point on the axis, which reach the image from different zones of the lens, strike the axis at different points; the distance between the point of intersection of the paraxial ray and that of a ray from any given zone of the lens measures the spherical aberration of that zone.

(ii.) *Coma.*—The magnification produced by the different zones of the lens is not a constant; hence the rays diverging from a point not on the axis of the lens are not brought to coincidence at one point in the image space; the intersection points of the rays from any given zone are distributed in a circle. The position of the centre and the diameter of this circle depends on the obliquity of the central ray of the pencil falling on the zone in question; hence the intersections of the rays from the object point and a plane through the image point are distributed over a fan-shaped area whose position and dimensions depend on the obliquity. This defect is known as Coma.

[1] See "Optical Calculations"; also "Telescope," §§ (1)-(3).
[2] See "Telescope," § (3).
[3] See "Lens Systems, Aberrations of."

(iii.) *Astigmatism.*—The rays of an oblique pencil diverging from an object point do not meet in one point in the image space; they do, however, pass through two "focal lines," lying approximately at right angles to each other. Of these the "primary" focal line is a short arc of a circle in a plane perpendicular to the axis of the lens, while the "secondary" line is more exactly a small figure of eight at right angles to the circular arc. The lens is astigmatic, and the nearest approach to point coincidence is a small circle, the "focal circle," lying between the two focal lines, and this is in many cases treated as the image of the object point.

(iv.) *Curvature.*—The focal circles corresponding to a series of object points lying in a plane perpendicular to the axis do not lie in a plane, but form a curved surface, cutting the axis at right angles. The image field corresponding to a plane object field is curved.

(v.) *Distortion.*—This is produced when the image field is a point-for-point representation of the object field, but on a scale which varies with the obliquity of the central rays of the pencil considered. Thus a straight line in the object field at right angles to the axis will be represented by a line in the image field, but this line will no longer be straight.

The above constitute von Seidel's five aberrations.

§ (12) Chromatic Aberration.—The position of the image corresponding to a given object point depends on the refractive index of the material of the lens, and this varies with the wave-length of the light used.

The images therefore formed by light of different wave-lengths will vary in size, and if white light is employed the resulting image will show coloured fringes. This defect is known as chromatic aberration. The various defects are fully discussed in the separate articles already mentioned, to which reference should be made.

For the general theory, "Optical Calculations."

For the telescope object glass, "Telescope."

For the microscope, "Microscope, Optics of the."

Levelling, Errors in. See "Surveying and Surveying Instruments," § (35).

Levelling and Levelling Instruments. See "Surveying and Surveying Instruments," §§ (29), (30); for precise levels, § (31).

Levelling Staves. See "Surveying and Surveying Instruments," § (32).

Levels. See "Spirit Levels."

Life-test (of Electric Lamps). See "Photometry and Illumination," § (77) *et sqq.*

LIGHT, ABSORPTION OF, IN COLOURED GLASSES, examined by R. Zsigmondy. See "Glasses, Coloured," § (4).

Tabulated results of R. Zsigmondy. See *ibid.* § (4).

LIGHT, DIFFRACTION OF

IT is known[1] that the effect at any point of a wave of light may be found by calculating the effect due to each element of the wave and taking the sum.

Now let the displacement[2] over a small area dS at a point Q (*Fig.* 1) of a wave front be given by the expression $A \sin (2\pi/\lambda)(vt)$

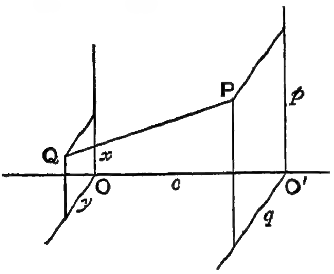


FIG. 1.

where t is the time, we proceed to find the displacement at a point P where QP is equal to r.

Since the intensity of the light, measured by its energy, diminishes as the square of the distance from the source and the energy of wave motion is proportional to the square of the amplitude, the amplitude will be inversely proportional to r. Moreover, since the disturbance at P has travelled from Q, a distance r, with speed v, the time occupied has been r/v. Thus the phase of the disturbance at P at time t will be the same as that at R, a time r/v previously. The amplitude is also clearly proportional to A, the amplitude at Q. Hence we have for the disturbance at P the expression

$$\frac{kA\,dS}{r} \sin \frac{2\pi}{\lambda}(vt - r),$$

k being a constant factor, and the whole effect at P is found by integrating this over the wave surface through Q. It will depend on the limits of this surface.

In this expression we have neglected the effect of the obliquity of QP to the directions of propagation and vibration at Q. We shall limit the discussion to the consideration of small areas about Q over which these angles do not vary much, and their effects may be considered constant and included in the factor k.

We assume that the front of the wave is parallel to the plane of the aperture. Take any point O in this plane as origin and two lines at right angles in the plane as axes of x and y respectively.

Consider a plane through P parallel to the wave front and let a line OO′ perpendicular to the wave meet this plane in O′, let OO′ be equal to c and let p, q be the distances of P from O′ measured parallel to the axes, then the co-ordinates of P are p, q, c, and we have

$$r^2 = PQ^2 = (p-x)^2 + (q-y)^2 + c^2.$$

Thus

$$r = c\left\{1 + \left(\frac{p-x}{c}\right)^2 + \left(\frac{q-y}{c}\right)^2\right\}^{\frac{1}{2}}$$

$$= c + \tfrac{1}{2}\frac{(p-x)^2 + (q-y)^2}{c} \text{ approximately.}$$

The result obtained by substituting this value of r in the expression for the disturbance for certain forms of the aperture can be evaluated in terms of integrals known as Fresnel's Integrals.[3]

We can, however, simplify the expressions by supposing the screen on which the effect is produced to be far from the aperture, so that p, q, and c are large compared with x and y.

In practice this is realised by placing a convex lens behind the aperture and observing the effect on a screen at the principal focus of the lens. We have as the value of r

$$r = c + \tfrac{1}{2}\frac{p^2 + q^2 - 2px - 2qy + x^2 + y^2}{c},$$

and on this assumption we can neglect $x^2 + y^2$ in comparison with the other quantities. Hence

$$r = c + \tfrac{1}{2}\frac{p^2 + q^2}{c} - \frac{2px}{c} - \frac{2qy}{c}$$

$$= f - \frac{2px}{c} - \frac{2qy}{c} \text{ say,}$$

where f is written for $c + \frac{1}{2}(p^2 + q^2)/c$. But p and q will in all cases be small compared with c, thus in the denominators we may replace c by f, and we have finally

$$r = f - \frac{2px}{f} - \frac{2qy}{f}.$$

Moreover, $dS = dxdy$, and the expression for the disturbance at P becomes

$$\frac{kA}{f}\iint dxdy \sin \frac{2\pi}{\lambda}\left\{vt - f + \frac{2px}{f} + \frac{2qy}{f}\right\}, \quad (1)$$

where f has been written for r in the denominator.

[1] See "Light, Propagation of," also "Light, Rectilinear Propagation of."

[2] Displacement is used as a general term, it may represent electric or magnetic force or may stand for an actual change of position of an electron, or of a particle of the ether.

[3] See Preston's *Light*, § 161.

(i.) *Rectangular Aperture.*—The expression given in (1) applies to an aperture of any form; thus for a rectangular aperture of sides $2a$, $2b$ parallel to the axes of x and y respectively the limits will be $-a$ to $+a$ and $-b$ to $+b$, and we have

Disturbance

$$=\frac{kA}{f}\int_{-a}^{+a}\int_{-b}^{+b}dxdy\sin\frac{2\pi}{\lambda}\left(vt-f+\frac{2px}{f}+\frac{2qy}{f}\right).$$

On integrating this we find for the intensity at P, represented by the square of the resultant amplitude, the expression

$$16a^2b^2\left\{\frac{f\lambda}{2\pi pa}\sin\frac{2\pi pa}{f\lambda}\right\}^2\left\{\frac{f\lambda}{2\pi qb}\sin\frac{2\pi qb}{f\lambda}\right\}^2,$$

or, as it may be written,

$$\bar{A}^2\times\left(\frac{\sin\theta}{\theta}\right)^2\left(\frac{\sin\phi}{\phi}\right)^2, \quad . \quad . \quad (2)$$

where $\bar{A}$ represents the area of the parallelogram and

$$\theta=\frac{2\pi pa}{f\lambda},\quad \phi=\frac{2\pi pb}{f\lambda}.$$

(ii.) *A Narrow Rectangular Slot.*—An important case occurs when one side of the aperture, say a, is very small in comparison with the other.

The angle θ is then always very small and the value of $\sin\theta/\theta$ is unity. The amplitude then varies as $\sin\phi/\phi$. This has a maximum when ϕ is zero, it is then equal to unity; it is a minimum—zero—for $\phi=\pi, 2\pi\ldots$ and a maximum for values between, which are given by the equation $\tan\phi=\phi$. These, it can be shown, are somewhat less than the odd multiples of $\pi/2$ excluding the first, but gradually approach them. The intensities of these maxima decrease somewhat rapidly. Thus the field consists of a bright band in the centre flanked on each side by a series of dark lines, equally spaced, with bright bands of rapidly decreasing brightness between; the distances between the maxima of the bright bands are unequal, but decrease as the distance from the central bright band increases.

(iii.) *Circular Aperture.*—Another case of importance when dealing with optical instruments occurs when the aperture is a circle. The diffraction pattern will be a series of circles having their centre at the point where a normal to the wave front through the centre of the aperture meets the screen and the distribution of the light will be the same along any radius of these circles. We can find it for the radius through the axis of x for which q is zero, p will then be the distance of the point from the centre of the diffracted rings. Thus we have

$$\text{Disturbance}=\frac{kA}{f}\int\int dxdy\sin\frac{2\pi}{\lambda}\left(vt-f-\frac{px}{f}\right), \quad (3)$$

the integrals being taken over a circle of radius a.

Let r, θ be the polar cos of $Q(x, y)$ relative to O.

Then $x=r\cos\theta$, $dxdy=rdrd\theta$; the limits for r are 0 to a, and for θ, 0 to 2π.

The disturbance

$$=\frac{kA}{f}\int_0^{2\pi}\int_0^a rdrd\theta\sin\frac{2\pi}{\lambda}\left(vt-f-\frac{pr\cos\theta}{f}\right)$$

$$=\frac{kA}{f}\sin\frac{2\pi}{\lambda}(vt-f)\int_0^a\int_0^{2\pi}rdrd\theta\cos\frac{2\pi}{\lambda}\left(\frac{pr\cos\theta}{f}\right),$$

for the term in the integral involving the quantity $\sin(2\pi/\lambda)(pr\cos\theta/f)$ clearly vanishes. The above expression can be integrated in a series,[1] and we find

Amplitude of disturbance

$$=\pi a^2\left\{1-\tfrac{1}{2}\left(\frac{m}{1!}\right)^2+\tfrac{1}{3}\left(\frac{m^2}{2!}\right)^2-\tfrac{1}{4}\left(\frac{m^3}{3!}\right)^2+\ldots\right\}, \quad (4)$$

where m is written for the value of

$$\frac{2\pi a}{\lambda}\cdot\frac{p}{f}.$$

The quantity p/f measures the sine of the angular radius of a ring as seen from the centre of the aperture, or since it is small we may take it as the angular radius. The series can be shown to be convergent. It has a first maximum when m is zero and passes through a series of maxima and minima as given in the following table taken from Verdel's *Optique Physique*:[2]

INTENSITY OF LIGHT DIFFRACTED BY A CIRCULAR APERTURE

Maxima.		Minima.	
m/π.	Intensity.	m/π.	Intensity.
0	1	0·610	0
0·819	0·01745	1·116	0
1·333	0·00415	1·619	0
1·847	0·00165	2·120	0
2·361	0·00078	2·621	0

If we call ϕ the angular deflection of the rings as seen from 0 we have approximately $m=(2\pi a/\lambda)\phi$; the diameter of the rings is 2ϕ, thus the diameters of the rings are given by the expression $(m/\pi)(\lambda/a)$, where m/π has the values given in the table; the width of the first dark ring is thus $0·610\ \lambda/a$. Hence the image of a point of light such as a star formed by the object glass of a telescope of aperture $2a$ is a disc having an angular breadth measured from the centre of the object glass of $0·610\lambda/a$. This disc is surrounded by a number

[1] See Preston's *Light*, § 163.
[2] See also *ibid.* § 163.

of rings of rapidly decreasing brightness with dark spaces between in which there is no light.

This assumes that the object glass is free from aberration, so that the wave from the distant point after passing the lens is accurately a sphere with the principal focus as centre.

LIGHT, DOUBLE REFRACTION OF

It is to Fresnel that we owe the earliest complete theory of double refraction applicable to both biaxial and uniaxial crystals, and though he based it on the hypothesis of the vibration of ether particles, the laws at which he arrived follow rigidly as a deduction from the electromagnetic theory.

If a particle in an isotropic elastic medium be displaced a short distance, the symmetry of the medium requires that the force called into play through the elasticity of the medium and which tends to bring the particle back to its equilibrium position should be in the direction of motion. The same is true if the displacement be electric or magnetic; the electric or magnetic force of restitution is proportional to the displacement and acts in that direction. It may be written as Kp, if p is the displacement and K a constant representing the electric—or magnetic—elasticity of the medium.

But in an anisotropic medium which has different properties in different directions this is no longer true; the force of restitution is not, in general, co-linear with the displacement.

In a crystal, however, there are three directions mutually at right angles along which the force of restitution is in the direction of displacement; these directions are known as the axes of the crystal. Now suppose that the forces produced by a displacement p along each of the axes respectively are a^2p, b^2p, and c^2p respectively. Consider a plane wave travelling through the crystal and construct an ellipsoid having its centre on the wave and its axes equal to $1/a$, $1/b$, and $1/c$ respectively. This ellipsoid was called by Fresnel the Ellipsoid of Elasticity. The wave front we are considering will cut the ellipsoid in an ellipse, and the direction of the two axes of this ellipse are found, both by theory and experiment, to be the only possible directions of vibration for light travelling in the direction in which the plane wave is being propagated.

In an isotropic medium the disturbance which constitutes the light can take place in any direction in the plane of the wave; in a crystal it is resolved into two in the direction OP, OQ of the two axes of this ellipse. Moreover, it can be shown that these two disturbances travel with different speeds; their velocities are given by 1/OP and 1/OQ respectively. The wave is thus split into two plane polarised waves, each travelling with the speed appropriate to its direction of vibration, and these velocities are, in each case, inversely proportional to the radius vector of Fresnel's ellipsoid drawn in the direction of vibration. Moreover, the possible directions of vibration for any plane wave are the axes of the section of Fresnel's ellipsoid by the wave. Thus a knowledge of Fresnel's ellipsoid enables us to calculate the velocity of wave propagation in all directions in the crystal. It is shown in treatises on electricity that, on the electromagnetic theory of light, the quantities a^2, b^2, c^2 are the three principal coefficients of inductive capacity of the medium, denoted usually by K_1, K_2, K_3. Again, it is known[1] how to determine the wave surface in any medium, crystalline or isotropic, when the velocity of wave propagation is known for all directions.

To describe the wave surface corresponding to an interval of time t we draw all possible wave fronts in the positions they would occupy at the instant t and describe the envelope of the series of planes thus obtained.

When this is done for a crystal in which the wave velocities are determined, as described above, a surface of somewhat complicated form is the result. The surface, known as Fresnel's wave surface, has two sheets, and its sections by the three principal planes of the crystal are in each case a circle and an ellipse. Assuming that a, b, c are in descending order

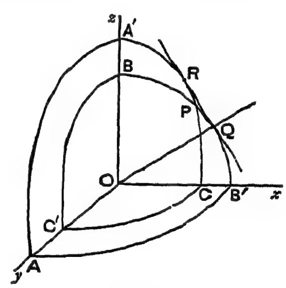


Fig. 1.

of magnitude, the sections will be as shown in *Fig.* 1.

In this figure Ox, Oy, Oz are the three crystallographic axes. Along them distances OA, OB, etc., are taken, such that

$$OA = OA' = a; \quad OB = OB' = b; \quad OC = OC' = c.$$

AA′, BB′, and CC′ are joined by circles of radii a, b, c respectively, while ellipses of semiaxes b and c, c and a, and a and b respectively join BC′, CA′, and AB′.

The circle BB′ cuts the ellipse CA′ in P,

[1] See article "Light, Propagation of."

and RQ is a common tangent to both circle and ellipse.

If we imagine the curve A′A to be turned about Oz so that A always lies on the ellipse AB′, the shape of the curve altering according to the law given by Fresnel until, ultimately, it coincides with A′PB′, it will trace out one octant of the outer sheet of Fresnel's wave surface.

Similarly, if BC′ rotates about the same axis, C′ remaining on the circle C′C, while the shape alters as required by Fresnel's laws until the curve coincides with BPC, it traces out the inner sheet of the wave surface.

Careful experiment [1] has verified to a high degree of accuracy the fact that this surface does represent the form of the wave in a biaxial crystal.

The quantities a, b, c measure respectively the velocities of waves in which the vibrations are parallel to the three axes respectively; thus the refractive indices corresponding to these waves are $1/a$, $1/b$, and $1/c$. If the mean refractive index $1/b$ differs less from the minimum $1/a$ than the maximum $1/c$ differs from the mean, *i.e.* if $1/b - 1/a$ is less than $1/c - 1/b$, the crystal is said to be positive; if, on the other hand, $1/b - 1/a$ is greater than $1/c - 1/b$, the crystal is negative.

According to the construction already explained, in order to find the position of the wave front which, after a time t, corresponds to a given wave front XY, we describe the wave surface for time t, taking any point in XY as centre, and draw a tangent plane parallel to XY.

It is clear from the shape of the surface that two such planes can, in general, be drawn, one touching each sheet; thus a given wave travelling through the crystal will break up into two; each of these will, as has already been pointed out, be polarised. The lines joining the centre of the surface to the points of contact of these two tangent planes are the rays corresponding to the two wave fronts.

But it is clear also, from the figure, that in the plane perpendicular to the mean axis in which the circle of radius b intersects the ellipse semiaxes a and b in the point P, there is a common tangent RQ to the ellipse and circle, and a plane through RQ at right angles to the plane xOz touches both the ellipse and circle. In the direction of the normal to this plane there will be only a single wave. The line OQ, which is at right angles to this wave, is known as an optic axis of the crystal. There will be a similar axis lying also in the plane zOx, but on the opposite side of Oz. Thus the crystal has two optic axes; light waves travelling along these are not doubly refracted. We can put this from another point of view, thus: The wave velocities are determined by the two axes of the elliptic section in which Fresnel's ellipsoid is cut by the plane of the wave. But an ellipsoid has two sections which are circular. If then the wave coincides with either of these, the two velocities of propagation become equal, each being equal to b, for this quantity can be proved to be the reciprocal of the radius of the circular section. Thus the two waves which, in general, are found, in this case coalesce and travel through the crystal with velocity b. The normals to these two sections are the two optic axes.

The optic axes lie in the plane containing the axes of greatest and least velocity. These axes bisect the angles between the optic axes, and are known as the acute bisectrix and the obtuse bisectrix of the angle between the optic axes.

A special significance attaches to the line OP joining O to P, the point of intersection of the circular and elliptic section of the wave surface. It is clear from the figure that two tangents can be drawn to the surface, touching it at the same point P; one of these touches the circle, the other the ellipse. The full investigation shows that, in reality, the surface has a tangent cone at P, and any plane touching this cone is also tangent to the surface. Now the ray corresponding to any wave plane is the line joining the centre to the point in which the plane touches the surface. All the planes, then, which touch the surface at P have the same ray and the same ray velocity. Thus the line OP is known as the axis of single ray velocity.

We arrive at the case of a uniaxial crystal by supposing two of the principal velocities to become equal; suppose, for example, that $a = b$; then Fresnel's ellipsoid becomes a spheroid, with axes $1/a$ and $1/c$, generated by the revolution of an ellipse about the axis Oz. The optic axes close up to this line. The section of this spheroid by any plane through its centre, other than the plane of xOy, is an ellipse, but one axis of each of these ellipses is equal to $1/a$; thus one wave of the two into which an incident wave is divided travels with constant velocity a and is refracted according to the ordinary law; the other suffers extraordinary refraction.

The wave surface becomes a sphere and a spheroid, the points A′B′ of *Fig.* 1 coincide, while the ellipse AB becomes a circle of radius a. The points Q, P, and R all coincide with A′ and B′. Thus Huyghens' construction, described earlier, follows as a deduction from Fresnel's more general theory. This also has been verified by experiment.[2]

[1] Glazebrook, "Plane Waves in a Biaxial Crystal," *Phil. Trans.*, 1878, p. 287.

[2] Glazebrook, "Double Refraction and Dispersion in Iceland Spar," *Phil. Trans.*, Part II., 1879, p. 421.

LIGHT FILTERS

§ (1) INTRODUCTION.—Light filters are used in many branches of physical work to modify the intensity or colour of the light passing through them. Either liquid filters in the form of solutions, or solid filters consisting of coloured glasses or stained films are commonly employed, though in some special cases gases have been used as filters.[1]

Liquid filters consist of solutions of coloured salts or dyes held in glass troughs with parallel sides, while coloured glasses and stained films of gelatine or collodion are used in a great variety of forms. The use of solutions has the advantage that they are very readily prepared and that, if they consist of inorganic salts, they are easily definable, so that standard filters can be prepared from a published description. To a lesser extent the same is true of dye solutions, but since dyes vary considerably in their purity, it is more difficult to state formulae for liquid dye filters in a form which can easily be duplicated, and solutions of dyes are often unstable, the dyes changing or precipitating. Such liquid filters are therefore often the cause of unsuspected errors.

If coloured glasses could be obtained in a great variety of absorptions, and could be reproduced exactly, they would undoubtedly be the most convenient for filters of all kinds. They are usually very stable to light, extremely convenient in use, not easily being damaged, and are cheap to produce, since it is only necessary to grind the two surfaces of the optical glass to optical planeness and to make the filter of a definite thickness. Unfortunately, it seems to be very difficult for glass makers to reproduce the colour of different meltings with any approach to exactness, and the number of coloured glasses available is quite limited, so that it is not possible to obtain in glass filters an approach to the range of colour which is available by the use of dyes.

The earlier stained filters were made by the use of collodion in which basic dyes were incorporated, but the difficulty of coating collodion accurately upon glass, as also the unsatisfactory nature of most of the basic dyes, has led to the almost complete replacement of collodion by gelatine as a vehicle for the acid dyes which are now used for the preparation of light filters. Modern gelatine light filters are therefore prepared by coating gelatine containing the necessary dyes upon plate glass, the coating being done by means of one of the coating machines commonly used for the preparation of photographic plates. The gelatine film, after drying, is stripped off the glass and can be used as a gelatine film filter or can be cemented between optical glasses of any required quality by means of Canada balsam.

On a small scale, filters are generally made by coating the glass to be cemented with the dyed gelatine and then cementing a cover-glass directly on to the dry gelatine film. This method is not advantageous on the large scale, since the coating of the large glasses is tedious and expensive; the coating is likely to vary in thickness from the centre to the edge owing to surface tension, and there is always danger that the drying gelatine in contracting will bend the glass, thus interfering with its optical properties. This latter difficulty can be mitigated to some extent by the addition of glycerine to the gelatine solution when coating the glass.

The colour of a filter is of course conditioned by its selective spectral absorption; thus a red filter absorbs the blue and green regions of the spectrum, a blue filter the red, yellow the blue, and a green both the red and the blue regions. This selective absorption is best expressed by means of the absorption curve of the filter, which can be measured by means of a spectrophotometer. The filter is placed in front of the instrument, and the light passing through the filter and also passing into the instrument without going through the filter is analysed in the spectroscope. A narrow region of the spectrum is isolated, and the absorption of the filter for each region of the spectrum in turn is measured photometrically. Then, if I be the intensity of the light passing through the filter, and I_0 that which has not undergone absorption, the transparency of the filter will be I/I_0, and this is usually expressed as a percentage if the transmission is required. From Lambert's law, however, $I=I_0e^{-kd}$.[2] This is more conveniently written using 10 as the logarithmic base, and for a light filter in which the thickness is constant it becomes

$$I=I_0 10^{-E},$$

where E is the absorption constant of the filter, and consequently

$$E=-\log\frac{I}{I_0}.$$

As a result of the Fechner law, absorption curves of filters expressed by plotting E against λ correspond more accurately to the appearance of the transmission as seen by the eye than those obtained by plotting the transparency, and these curves are used throughout the literature dealing with light filters.

[1] Thus bromine gas has been used as a filter for the ultra-violet by R. W. Wood and Plotnikoff.

[2] It seems to be customary to refer to this as "Lambert's law," but the originator of it was Bouguer (see his *Essai d'optique*, 1729), and the law of the absorption of light should certainly be called "Bouguer's law."

Following Potapenko,[1] light filters are classified preferably according to their absorption spectrum and not according to the purpose for which they are to be used, since if this latter course is taken the same filter will be classified in many different ways; thus the red filter suitable for three-colour photography is also used as a contrast filter in commercial photography, as a photomicrographic filter, and in spectroscopy, while there are undoubtedly many other uses to which it can be put. Classifying filters, therefore, according to their spectral absorption, we may divide them into:

(i.) Selective filters, which transmit only a selected region of the spectrum, more or less narrow.

(ii.) Compensating filters, which have a more gradual absorption, and which transmit in

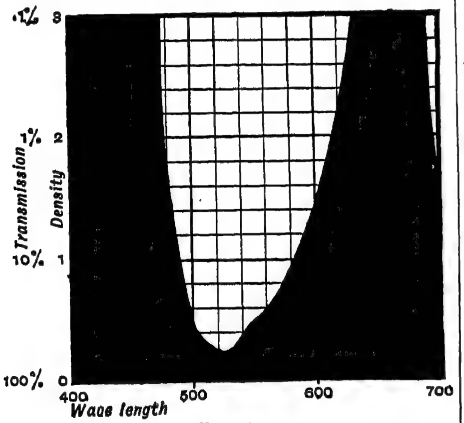


FIG. 1.

greater or less intensity practically the whole spectrum.

(iii.) Subtractive filters, which remove only a small portion of the spectrum, transmitting the remainder, so that for every selective filter we can theoretically have a corresponding subtractive filter.

The selective and subtractive filters will therefore have a comparatively sharp absorption curve, while compensating filters will generally be gradual in their absorption. Monochromatic filters (which is the title used by Potapenko for the filters here called selective) are those which transmit only a narrow region of the spectrum, so narrow that the term *monochromatic* is not entirely inappropriate, though it must be understood that no filter can be strictly monochromatic in the true sense of the word unless used with a linear emission spectrum. An example of a selective filter is shown in *Fig.* 1, which is the absorption spectrum of a typical green filter used in three-colour photography. *Fig.* 2 shows a narrow banded monochromatic filter;

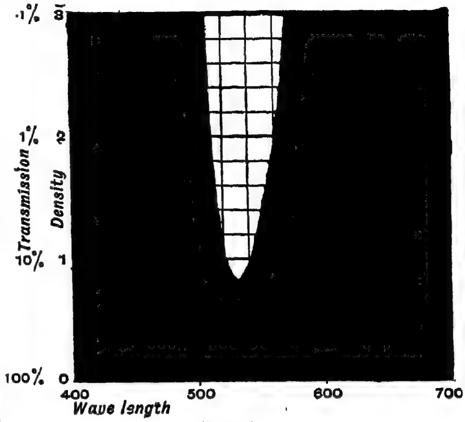


FIG. 2.

Fig. 3 a compensating filter used for compensating the spectrum of a vacuum tungsten lamp to balance daylight; and *Fig.* 4 is the subtractive filter which is complementary to the green filter shown in *Fig.* 1.

§ (2) APPLICATION OF LIGHT FILTERS — PHOTOGRAPHY.—Filters used in photography may be classified according to their use under the heads of:

(i.) *Orthochromatic filters*, used to correct the selective sensitiveness of photographic materials

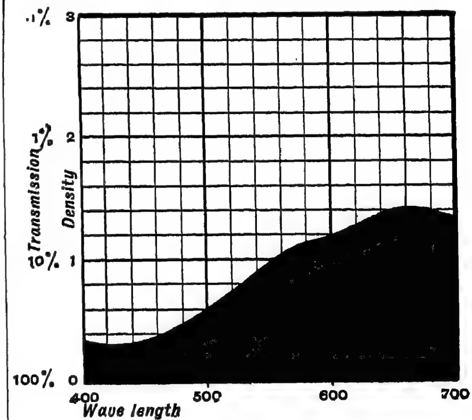


FIG. 3.

in order to give a reproduction of the brightness of coloured objects more closely approximating that perceived by the eye.

(ii.) *Contrast filters*, used selectively to modify the brightness of some special colour or group of colours.

[1] G. V. Potapenko, *Journal Russian Physical and Chemical Society*, 1916, xlviii. 790; English translation, *British Journal of Photography*, 1921, lxviii. 507. This is by far the fullest and best article published on light filters.

(iii.) *Selective filters*, for two- or three-colour photography, with their corresponding synthetic projection filters if the processes of colour photography used are of the additive type.

(iv.) *Compensating filters*, used to adjust the light entering a photographic system in order

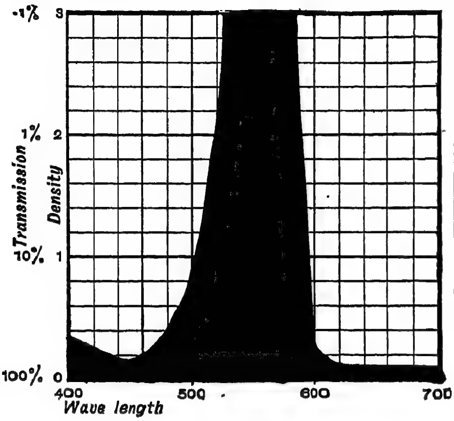


FIG. 4.

to fulfil some particular condition required in colour photography.

(i.) *Orthochromatic Filters.*—Since photographic plates are more strongly sensitive to blue light than to any other spectral region, and since the maximum of the visual sensitiveness is in the yellow-green, and the blue and blue-violet are comparatively dark colours to the eye, orthochromatic filters are yellow in colour and absorb the ultra-violet, violet, and to a lesser extent the blue. The orthochromatic filters used in photography often do not perform their task completely, since the removal of the light of short wave-lengths naturally increases the exposure, and it is often impossible to give sufficient exposure to use the filter required for complete correction. It is very common, therefore, in photography to compromise by using a filter which removes most of the violet and ultra-violet but does not fully correct the uneven sensitiveness of the plate.

The earlier orthochromatic filters used by photographers were made of glass which was coloured by carbon in the pot, and which was brown rather than yellow, transmitting a large amount of ultra-violet light; and even when dye filters were used, dyes which transmitted a large amount of ultra-violet light were often employed through lack of sufficiently rigorous spectroscopic examination.

The number of dyes suitable for the preparation of orthochromatic filters is very limited. Of those giving a sharp absorption in the violet, with a complete absorption of the ultra-violet, picric acid and the picrates are the most satisfactory, but unfortunately they are very unstable and can scarcely be used in practice. Until 1906 the best dye available was Tartrazine, but this has a very considerable transmission in the ultra-violet and was replaced for filter-making in that year by a dye manufactured by the Hoechst Dye Works under the name of "Filter Yellow." During the war, a dye having almost as great stability as Filter Yellow with a considerably sharper absorption was prepared from phenyl glucosazone and was used for light filters, especially for aerial photography, under the title of "Eastman Yellow" (*Fig.* 5).

In order to enable landscapes to be taken in which the clouds are well rendered without unduly increasing the exposure, graded filters have been employed, these being placed in front of the lens in such a way that the strongly absorbing region of the filter corresponds to the sky and the weaker absorption to the foreground. Filters are sometimes made having the upper half coloured and the lower part clear, which seem to be as satisfactory in practice as the graded filters, provided they are not used too far from the lens. A complete

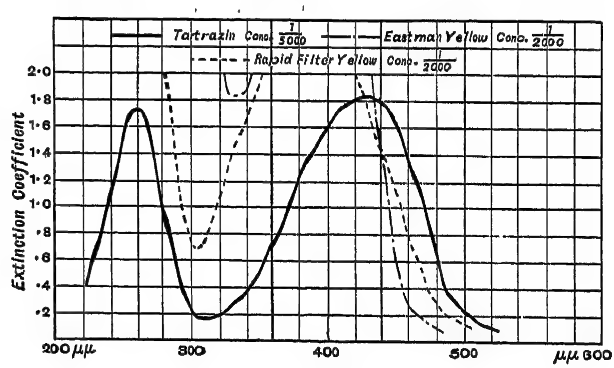


FIG. 5.

discussion of the use of graded filters will be found in a paper by J. Weissmann.[1]

(ii.) *Contrast Filters.*—The preparation of contrast filters is in many ways easier than that of orthochromatic filters, since many dyes are available for the preparation of red or green filters. The greatest difficulty arises when it is necessary to remove the extreme red from a green or blue filter. Only one blue

[1] *Photographische Rundschau*, 1913, p. 182.

dye and three green dyes are known which have not a sharp transmission band at the red end of the visible spectrum. These are Toluidine Blue, which has no transmission band in the red of shorter wave-length than 750 $\mu\mu$; Anthraquinone Green, which is the green dye corresponding to Toluidine Blue; Filter Blue-Green, of which the composition is unknown, but which extends in absorption to at least 800 $\mu\mu$ and at the same time has greater transmission in the green than Anthraquinone Green; and Naphthol Green, which absorbs at least as strongly as Filter Blue-Green, but which unfortunately has a great deal of general absorption, especially in the blue-green and blue regions of the spectrum, so that it is impossible to use it in blue filters, and it greatly lessens the transmission of green filters. Since the first three of these dyes have only been introduced in the last twelve years, the older blue and green light filters all show a very considerable transmission of red light. Attempts have been made to diminish this by the use of Methylene Blue, but unfortunately this dye is unstable and is rapidly destroyed by heat, so that it is not at all suitable for use in light filters.

A set of contrast filters which will satisfy almost all photographic requirements will include a strong yellow filter cutting at about 500 $\mu\mu$; an orange filter cutting at 570 $\mu\mu$; a red cutting at 590 $\mu\mu$; and a deep red cutting at 610 $\mu\mu$, in addition to the orthochromatic filters and to green and blue filters which can conveniently be those used also for tricolour photography.

(iii.) *Selective filters* used in colour photography will include for ordinary work a set of red, green, and blue filters so adjusted that they just overlap and that they give approximately even exposures when used with daylight upon the most highly sensitive panchromatic plates available. For colour cinematography a somewhat lighter set of filters may be used, though it is not possible to lighten the red filter very much, a filter cutting at 575 $\mu\mu$ being the lightest filter that can be used to give at all satisfactory rendering.

For two-colour photography as deep a red filter is required as for three-colour work. The exact shade of the two-colour green filter has been a matter of some discussion and a good deal of experiment. Most workers use the tricolour green, though a somewhat bluer green has often been advocated. The disadvantage of too blue a green is that foliage and grass tend to be reproduced as brown, and for taking filters for two-colour work the tricolour red and green filters would seem on the whole to be the most suitable.

In the additive projection processes the filters must be adjusted to the light source used so that the screen is white. For most work the same red and green filters that are used for taking will be satisfactory for tricolour projection, a lighter blue being used according to the colour of the illuminant employed. For two-colour projection filters a much bluer green than that used for taking is naturally required, and the filter should be rather a blue-green than a green.

(iv.) *Compensating filters* are of special importance in connection with the screen plate processes of colour photography. In these processes the three-colour unit filters of the screen must be adjusted in area and colour so that the screen itself is neutral when looked at by a white light, while the emulsion sensitiveness, the colour of the screen units, and the compensating filter must be so adjusted that greys are reproduced as greys in the finished picture. Normally, the emulsion has predominating blue sensitiveness, and the compensating filter must therefore be yellow, though not infrequently a brownish or pinkish filter is required. Such compensating filters are usually adjusted by trial and error.

§ (3) PHOTOMETRY.—When comparing light sources of different colours upon the photometer the colour difference introduces difficulty in making an accurate balance. This difficulty can be much lessened by the use of a colour filter adjusted to equalise the colour in the two fields of the photometer, the transmission of the filter being first determined by a set of observations. Such photometric filters will be of two types: yellowish filters for reducing high efficiency lamps to the colour corresponding to lower efficiencies, and bluish filters for the inverse process. The absorptions of such filters must naturally be gradual, and they should not display any very marked spectral bands or rapid changes in their spectral curve.

§ (4) MICROSCOPY.—For photomicrography filters are employed especially in the photography of stained sections or other objects displaying strong colour. Thus, if any colour is to be rendered as dark as possible, it must be viewed or photographed by light which is absorbed by the colour, that is, by light of the wave-lengths contained within its absorption band. On the other hand, when a subject is of a uniform colour, in order to render detail it must be examined by the light which it transmits; thus, if the usual insect preparations, which are strongly yellow, are photographed by blue light as on an ordinary plate, a black detailless mass is shown in contrast to the background, while the use of a strong yellow or red filter will enable the whole of the detail of the structure to be photographed. Filters for microscopy are arranged so that they can be used in pairs, the spectrum being thus divided into monochromatic portions. By the examination of any specimen through

these filters singly and in pairs the best choice of the filters for use can be made. Light filters are also valuable for visual use with the microscope. Thus, blue and green filters are convenient for modification of the intensity of the light and for the restriction of the spectral region used in visual work. A deep blue filter is especially useful where the highest resolving power visually possible is required, as in the visual examination of diatom structures. Yellow, orange, and red contrast filters are convenient in the observation of detail in insect mounts or as contrast filters for stained preparations.

§ (5) Spectroscopy, Photo-chemistry, etc.—Light filters are often used in spectroscopy for the elimination of portions of the spectrum, such as the absorption of other orders when using a grating spectroscope. They are also of particular service in spectro-photometry for the elimination of scattered light. When using an absorption spectro-photometer for the measurement, for instance, of a solution of bichromate, suppose that the slit is set just inside the absorption band; then, one of the spectra is darkened throughout its length by the equalising apparatus, while in the other spectrum the red, orange, and yellow are entirely undarkened. The light from these regions of the spectrum will be scattered throughout the field and will change the apparent colour of the green or blue light transmitted by the solution at the edge of the absorption band. Satisfactory results in the use of the instrument can be obtained by the employment of filters which will absorb the whole of the red and yellow light, transmitting only the region in which measurements are required.

Monochromatic filters are often of use in spectroscopy and in general physical work, but unfortunately it is not possible as a general rule to get monochromatic filters of great efficiency. The transmission of most monochromatic filters is only 10 per cent or less if the filters are really narrow, and this loss of light seriously limits their application. There is one monochromatic filter, however, which is very efficient. This is the filter resulting from the employment of didymium salts, which have an extremely sharp absorption band between 550 $\mu\mu$ and 580 $\mu\mu$, thus cutting out very cleanly the mercury yellow lines. By the combination of this filter with a strong yellow filter, the green line of the mercury lamp can be obtained in high intensity and almost entirely free from other radiation, thus supplying the strongest monochromatic source of light known. This is of particular advantage in work with the interferometer.

§ (6) Optical Properties.—The optical effects produced by the use of colour filters in front of lenses may be divided into two classes: the first includes those resulting from the use of a theoretically perfect filter; the second, those which result from imperfections in the filters, that is, from departures from conditions of plane parallelism in the faces and of equality in thickness between the filters of a set. Computation shows that the effects resulting from the use of perfect filters are too small to be of any practical importance, the effect both on spherical aberration and on curvature of field being well within the limit of tolerance, while the effect of the dispersion on the difference in image size between different colour filters of a set is also negligible. The effects upon the definition and size of image introduced by the use of filters which are imperfect are very much more complicated than those which result from the use of perfect filters. Since the errors of surface may be of any type, the investigation of the effects produced can only be performed experimentally. The most convenient testing apparatus consists of a telescope objective of about four feet focal length on which the filter is placed, the image of a test object being examined by means of an eyepiece.[1]

C. E. K. M.

LIGHT, INTERFERENCE OF

§ (1) Introduction.—The problem of the interference of light and its historical development are closely bound up with the question of the nature of light. Two theories of light have held sway at different periods in the history of optical science, namely, the wave theory and the corpuscular or emission theory. According to the former, which was first definitely propounded by Huygens in 1678, light is a wave motion propagated in a continuous hypothetical medium called the ether, which is supposed to pervade all space. Although Huygens regarded the waves as longitudinal, he was able to explain a number of phenomena, such as reflection and refraction, but he could not account for the rectilinear propagation of light. For this reason the theory was not accepted by Newton, who elaborated the corpuscular or emission theory, according to which luminous bodies emit material corpuscles which travel along straight paths, the sensation of vision being produced by the mechanical impact of these corpuscles on the retina. This theory leads to the correct law of refraction, but it also leads to the result that the velocity of light should be greater in a dense medium than in a less dense one. The opposite result follows from the wave theory and has been experimentally proved to be correct; this forms a crucial

[1] Cp. C. E. K. Mees, *British Journal of Photography*, 1917, p. 462.

test of the two theories. The fact that the emission theory was not finally overthrown until the beginning of the nineteenth century, when interference was explained on the basis of the wave theory, is largely due to its being held by such a great authority as Newton.

§ (2) MATHEMATICAL ANALYSIS.—The interference of light may, according to the wave theory, be expressed in mathematical form as follows.

The general differential equations of wave motion in a homogeneous medium are given by

$$\frac{\partial^2\xi}{\partial t^2}=V^2\left(\frac{\partial^2\xi}{\partial x^2}+\frac{\partial^2\xi}{\partial y^2}+\frac{\partial^2\xi}{\partial z^2}\right)$$

and similar equations in η and ζ, where t is the time, (x, y, z) are rectangular co-ordinates, V is the velocity of propagation of the wave motion, and (ξ, η, ζ) are, according to the elastic wave theory, the components of the displacement of an ether particle from its position of rest or, according to the electromagnetic theory, the components of the electric or magnetic field strengths.

When the disturbance travels in the direction of the x axis this reduces to

$$\frac{\partial^2\xi}{\partial t^2}=V^2\frac{\partial^2\xi}{\partial x^2}.$$

Confining our attention to the case of a wave propagated along the direction of the x axis and writing λ for $V\tau$, we find that the disturbance in the wave may be represented by the equation

$$\xi=A\cos\frac{2\pi}{\tau}\left(t-\frac{x}{V}\right)=A\cos 2\pi\left(\frac{t}{\tau}-\frac{x}{\lambda}\right).$$

This represents a wave, for at a given point —for which x is constant—the motion repeats itself at times t, $t+\tau$, $t+2\tau$, etc., and at a given time—for which t is constant—it repeats itself at the points x, $x+\lambda$, $x+2\lambda$, etc., or x, $x+V\tau$, $x+2V\tau$, etc. Thus it can be represented by a cosine curve, repeating itself at lengths λ, 2λ, . . . which moves forward with speed V.

A is the amplitude, τ the period of vibration, and λ the wave-length of the vibratory motion of the ether particle. The intensity of such a light wave is proportional to A^2, for it is proportional to the kinetic energy of the vibrating ether particle, and this in turn varies as the square of the velocity of vibration; the velocity is given by the equation

$$\frac{\partial\xi}{\partial t}=-A\frac{2\pi}{\tau}\sin\left(t-\frac{x}{V}\right),$$

and the average value of its square is proportional to A^2.

We may now consider the effect of two light waves of the same period propagated in a homogeneous medium, the vibrations being supposed parallel. They may be represented by

$$\xi_1=A_1\cos 2\pi\left(\frac{t}{\tau}-\frac{x}{\lambda}\right)$$

and $$\xi_2=A_2\cos 2\pi\left(\frac{t}{\tau}-\frac{x+\delta}{\lambda}\right),$$

where A_1, A_2 are the amplitudes of the two motions, and $2\pi\delta/\lambda$ is the phase difference, that is, the crests of the one wave are always a distance δ ahead of those of the other. Now, according to the principle of the superposition of wave motion, the resultant motion of a particle simultaneously subjected to a number of different displacements is obtained by summing up the components of these displacements. Thus in the case of the two given light waves the resultant displacement is given by

$$\xi=\xi_1+\xi_2=\left(A_1+A_2\cos 2\pi\frac{\delta}{\lambda}\right)\cos 2\pi\left(\frac{t}{\tau}-\frac{x}{\lambda}\right)$$
$$+A_2\sin 2\pi\frac{\delta}{\lambda}\sin 2\pi\left(\frac{t}{\tau}-\frac{x}{\lambda}\right).$$

Now let

$$A_1+A_2\cos 2\pi\frac{\delta}{\lambda}=A\cos 2\pi\frac{\Delta}{\lambda}$$

and $$A_2\sin 2\pi\frac{\delta}{\lambda}=A\sin 2\pi\frac{\Delta}{\lambda}.$$

Then $$\xi=A\cos 2\pi\left(\frac{t}{\tau}-\frac{x+\Delta}{\lambda}\right),$$

where

$$A^2=A_1^2+A_2^2+2A_1A_2\cos 2\pi\frac{\delta}{\lambda}.$$

The resultant motion is thus given by a wave which has the same period as the original wave, but a different amplitude, and is in a different phase. If the original phase difference $2\pi(\delta/\lambda)=2n\pi$, that is, if $\delta=2n(\lambda/2)$, where n is a positive integer, A has its maximum value, A_1+A_2. On the other hand, if $2\pi(\delta/\lambda)=(2n+1)\pi$, that is, if $\delta=(2n+1)(\lambda/2)$, A has its minimum value, A_1-A_2. If δ is an odd number of quarter wave-lengths,

$$\xi_2=A_2\sin 2\pi\left(\frac{t}{\tau}-\frac{x}{\lambda}\right),$$

and $A^2=A_1^2+A_2^2$, that is, the total intensity is equal to the sum of the intensities of the two waves. If $A_1=A_2$ the total intensity is four times that of each wave when $\delta=2n(\lambda/2)$, and is zero, that is, the waves interfere, when $\delta=(2n+1)(\lambda/2)$. This does not mean that any energy is destroyed; what happens is that the distribution of the energy is modified.

The treatment may be extended to the case of any number of waves, but the brief sketch just given serves to illustrate the method whereby the phenomenon of interference may be explained by means of the wave theory of light.

§ (3) Young's Experiment.—True interference of light was first observed by Young; the experimental arrangement he employed was described in his lectures which were published in 1807. Light from a slit was allowed to fall on a screen in which there were two small pinholes placed close together. Interference bands were then observed on another screen placed in a position where the rays, after being diffracted at the pinholes, overlapped. The phenomenon is really one of interference of diffracted light,[1] because the rays which pass directly through the pinholes will not overlap. It will be of interest, however, to consider the explanation of the phenomenon by means of the wave theory on the assumption that the two pinholes do act as sources of light in which the vibrations bear a constant phase relationship to each other, for, as will be seen later, it is possible to obtain the necessary conditions by optical means. Let S_1, S_2 (*Fig.* 1) represent two

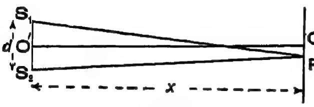


Fig. 1.

pinholes acting as sources of light. If we assume that the ether particles at S_1 and S_2 are vibrating parallel to each other in the same phase, and that the waves being sent out have the same amplitude and the same period, we may consider the effect at a point P on a screen OP, set parallel to the line joining S_1S_2, at a position where the beams from S_1 and S_2 overlap. Draw $O'O$ through the mid-point O' of S_1S_2 perpendicular to the screen. Let $S_1S_2=d$, $O'O=x$, and $OP=y$. Then

$$S_1P^2=x^2+\left(\frac{d}{2}+y\right)^2, \quad \text{and } S_2P^2=x^2+\left(\frac{d}{2}-y\right)^2.$$

Therefore $\quad S_1P^2-S_2P^2=2yd.$

If d and y are each small in comparison with x, we have

$$S_1P+S_2P=2x.$$

Thus $\quad S_1P-S_2P=\frac{yd}{x}.$

If the difference S_1P-S_2P is equal to an even number of half wave-lengths, that is $2n(\lambda/2)$, where n is an integer and λ is the wave-length, the waves will reinforce each other; there will therefore be a maximum intensity at P when

$$y=\frac{2nx\lambda}{2d}.$$

If, however, the difference S_1P-S_2P is equal to an odd number of half wave-lengths, that is $(2n+1)(\lambda/2)$, the waves will interfere; there will therefore be darkness at P when

$$y=\frac{(2n+1)x\lambda}{2d}.$$

It may easily be shown that, if a line is drawn on the screen so as to pass through P and be perpendicular to OP, the total disturbance at a point P′ on this line is the same as at P, if PP′ is small. Thus the appearance on the screen will be that of a series of parallel alternating bright and dark bands,[2] the distance between the centres of two consecutive bright bands or dark bands being given by

$$D=\frac{x\lambda}{d}.$$

The central band of the system, that is the one at O, is bright.

The positions of these bands or fringes depend on the wave-length of the light used. Thus, if white light is employed, there will be an infinite series of bright and dark bands corresponding to the infinite range of wave-lengths in white light. The result will be that only a few bands near the centre will be at all clear, the outer ones overlapping each other; the central band will be white. Thus, in order to obtain a considerable number of bands, monochromatic light must be employed.[3]

The above treatment will hold for all cases where the interfering beams come from two real or virtual images of one and the same luminous point. This requirement is fundamental for the production of interference fringes, it being impossible to obtain interference between beams which come from two entirely independent sources. The reason for this is that the phase of the disturbance from a given source, even if of constant wave-length, alters frequently and irregularly, so that there can be no fixed relationship between two independent sources. If, however, the two sources are real or virtual images of the same source, there will be a constant phase relationship between the vibrations in the two and interference systems will be observable. In such a case the source must have been maintained in regular vibration for millions of periods.

It is interesting in this connection to note that the number of fringes which can be observed gives us information with regard to the number of regular vibrations performed

[1] Diffraction phenomena, which may be looked upon as a special class of interference phenomena depending on the restriction of apertures, will not be discussed in this article.

[2] The bands are only approximately straight; actually they are the sections of a system of hyperboloids of revolution round the line S_1S_2 having S_1 and S_2 as foci.

[3] See, however, later on under "Achromatic Fringes," § (5).

by each molecule in the luminous source before a sudden change in phase takes place, for at the nth dark fringe interference takes place between vibrations which left the two sources at times $n\tau$ apart, where τ is the period of vibration. In the case of sodium light Fizeau counted about 50,000 fringes, while in more recent times path differences corresponding to as many as a million fringes have been obtained in the case of certain kinds of monochromatic light.

§ (4) Methods of producing Interference Fringes.—A great number of methods of producing interference fringes of the type just discussed have been devised, but they all practically depend on the formation, by means of some optical system, of two images (real or virtual) of a narrow source of light, such as an illuminated slit. These images then act as secondary sources, fringes being observed in the region where the beams of light which come from these sources overlap. In each case the distance D between consecutive bright or dark bands is given by

$$D=\frac{x\lambda}{d},$$

where x = distance between sources and plane in which the bands are observed,
d = distance between the secondary sources,
λ = wave-length.

Some of the more important optical systems devised for the purpose mentioned may be described as follows.

(i.) *Fresnel's Two-mirror System.* — This is one of the systems that Fresnel [1] suggested for the purpose of overcoming the objection that diffraction plays a part in Young's experiment. A plan of the arrangement is shown in *Fig.* 2.

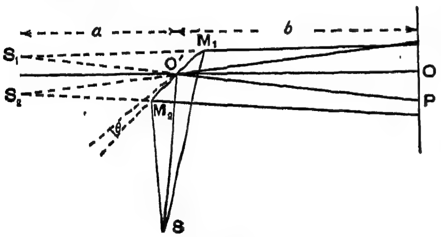


Fig. 2.

The light from a slit S is reflected from two plane mirrors $O'M_1$, $O'M_2$, the normals to which are inclined at a small angle θ, the line of intersection of the mirrors being parallel to the length of the slit. The reflected beams of light appear to come from S_1 and S_2, the images of S in the two mirrors, and interference bands parallel to the slit are obtained on a screen OP, where the two beams overlap. Now let $O'S=a$ and $O'O=b$. The angle subtended by S_1S_2 at O′ is equal to twice the angle between the mirrors. Also S_1 and S_2 are situated as far behind the mirrors as S is in front of them. Thus it follows that $O'S_1=O'S_2=O'S=a$, and $S_1S_2=2a\theta$.

Hence $$D=\frac{(a+b)\lambda}{2a\theta}.$$

The Fresnel mirror system may be modified by the introduction of a lens to form two real images of the slit.[2] This is the arrangement which Righi [3] used in his experiment for demonstrating the existence of light-beats.

(ii.) *Vautier's Three-mirror System.*—In this system [4] three equidistant mirrors are set up approximately parallel to each other, the inner one being silvered on both sides and the others on their inner sides only. The two interfering beams of light are formed by reflection first at each of the two outer mirrors and then at the central mirror. If the mirrors were set accurately parallel to each other the beams would be contiguous and no interference would take place; one of the outer mirrors is therefore set up so as to make a small angle with the others. Vautier has also employed an arrangement whereby the light incident on the mirrors is rendered parallel by the introduction of a lens; the beams after reflection then pass through another lens which forms two real images of the source of light.

(iii.) *Michelson's Mirrors.*—Michelson [5] forms two virtual images of a source by double reflection at two mirrors which are inclined to one another at an angle which is very little less than 90°. The arrangement has the advantage over that of Fresnel that it does not require such accurate adjustment of the mirrors.

(iv.) *Fresnel's Bi-prism.* — Fresnel [6] also obtained interference fringes by making use of refraction instead of reflection to form two adjacent virtual images of a source. The light from a slit S (*Fig.* 3) is refracted by the

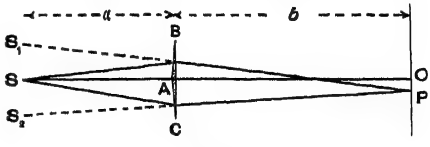


Fig. 3.

bi-prism ABC, the edge A of the obtuse angle dividing the light into two beams which, after

[1] A. Fresnel, *Œuvr. compl.* i. 150, 186; ii. 17, 52.

[2] Cf. E. Mascart, *Traité d'optique*, 1889, i. 180.
[3] A. Righi, *Journ. de Physique*, 1883, ii. 442.
[4] G. Vautier, *Comptes Rendus*, 1903, cxxxvii. 615; *Journ. de Physique*, 1903 (4), ii. 888.
[5] A. A. Michelson, *Am. J. Sci.*, 1890 (3), xxxix. 216.
[6] A. Fresnel, *Œuvr. compl.* i. 330.

refraction, appear to come from S_1 and S_2. Interference fringes can then be observed in the region where the two beams overlap, provided that the edge A is accurately parallel to the slit S. Let δ be the deviation produced by each half of the prism, μ the refractive index of the prism, and $\alpha = \angle ABC = \angle ACB$. Then

$$S_1S_2 = 2a \sin \delta = 2a(\mu - 1)\alpha,$$

if α is very small, where $a = SA$.

Thus the bi-prism is equivalent to a pair of Fresnel mirrors inclined at an angle $(\mu - 1)\alpha$. The distance between two consecutive bright or dark bands on a screen OP at a distance b from the bi-prism is then given by

$$D = \frac{(a+b)\lambda}{2a(\mu - 1)\alpha}.$$

If white light is used, there is greater overlapping of the bands due to different colours than with Fresnel's mirrors, because in the expression for D a decrease in the value of λ is accompanied by an increase in the value of μ, owing to the dispersion caused by the prism.

A simple and very useful method of measuring S_1S_2 has been suggested by Glazebrook.[1] A convex lens is inserted between the bi-prism and the microscope which is used for measuring the separation of the bands, and is adjusted so as to form images of S_1 and S_2 in the focal plane of the microscope. The distance d_1 between these images is measured. The lens is then moved into the second position, in which it forms images of S_1 and S_2 in the focal plane of the microscope, the distance d_2 between these images being measured. Since the magnification in the one position is the reciprocal of the magnification in the second position, it follows that

$$S_1S_2 = \sqrt{d_1 d_2}.$$

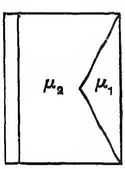


FIG. 4.

A modified form of Fresnel bi-prism, due to Winkelmann,[2] is illustrated in *Fig.* 4. The bi-prism, of refractive index μ_1, is mounted in a cell containing a liquid, such as benzene, of refractive index μ_2, the cell being covered by a plane parallel piece of glass. In this case

$$D = \frac{(a+b)\lambda}{2a(\mu_1 - \mu_2)\alpha},$$

where the symbols have the same significance as before. The advantage of this arrangement is that the obtuse angle of the bi-prism may be made much less obtuse than in the case of the bi-prism used in air.

[1] R. T. Glazebrook, *Physical Optics*.
[2] A. Winkelmann, *Zeits. Instrumentenk.*, 1902, xxii. 275.

(v.) *Bi-plates.*—Another method of obtaining two adjacent virtual sources for producing interference fringes is to use a bi-plate,[3] which consists of two parallel plates of glass of equal thickness (preferably cut from the same piece of glass) inclined to one another (*Fig.* 5).

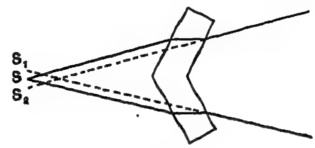


FIG. 5.

The beams of light after refraction appear to come from the virtual images S_1 and S_2. Two real images of S_1 and S_2 are formed by the introduction of a lens. Now the lateral displacement of a ray in passing through a parallel plate of thickness t is given by

$$\frac{t \sin (i - r)}{\cos r},$$

where i and r are the angles of incidence and refraction respectively. Hence

$$S_1S_2 = \frac{2t \sin (i - r)}{\cos r} = 2t \sin i \{1 - \cot i \tan r\}.$$

If the plates are inclined at angle 2θ we have $i = 90° - \theta$ approximately. Therefore

$$\sin i = \cos \theta, \quad \cos i = \sin \theta, \quad \cot i = \tan \theta,$$

$$\sin r = \frac{\cos \theta}{\mu}, \quad \cos r = \frac{\sqrt{\mu^2 - \cos^2 \theta}}{\mu},$$

$$\tan r = \frac{\cos \theta}{\sqrt{\mu^2 - \cos^2 \theta}},$$

where μ is the refractive index of the glass. Thus

$$S_1S_2 = 2t \cos \theta \left\{1 - \frac{\sin \theta}{\sqrt{\mu^2 - \cos^2 \theta}}\right\},$$

and the value of D may then be deduced.

(vi.) *Billet's Split Lens.*—In this system[4] (*Fig.* 6) two real images, S_1 and S_2, of an

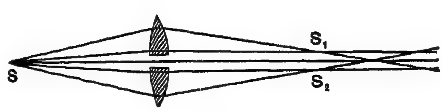


FIG. 6.

illuminated source are formed by two halves of a lens which can be separated or brought close together by means of a micrometer screw, the motion being along a line perpendicular to the optical axis.

(vii.) *Meslin's Split Lens.*—This is a modification of the previous arrangement, the

[3] J. Jamin, *Cours de physique*, 1866 (3), 524.
[4] F. Billet, *Ann. Chim. et Phys.*, 1862 (3), lxiv. 386.

two halves being relatively displaced along the optical axis.[1] The two real images, S_1 and S_2 (*Fig.* 7), are then formed on the optical axis and interference fringes can be observed in the region between S_1 and S_2. The points of equal phase difference in this case lie on ellipsoids of revolution, having S_1 and S_2 as

FIG. 7.

foci. A system of semicircular[2] fringes is then obtained. In both this and the previous case the distance S_1S_2 can be calculated from the separation and the optical constants of the lens halves.

(viii.) *Fresnel's Three-mirror System.*—In this system three mirrors—M_1, M_2, and M_3 (*Fig.* 8)—are employed. One of the secondary

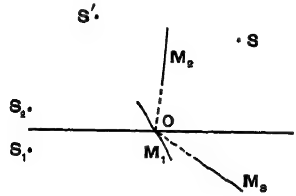


FIG. 8.

sources is the image S_1 of S in the central mirror M_1. The other secondary source S_2 is the image in M_3 of S′, which in turn is the image of S formed in M_2. Thus one of the beams is once reflected at M_1, while the other is twice reflected at M_2 and M_3. The mirrors M_2 and M_3 are set up in such a position that their planes intersect at O, the centre of M_1, making approximately equal angles with the plane of M_1. The central band of the interference system is black, from which Fresnel concluded, as Young had done in order to explain the black centre in Newton's rings, that there is a loss of optical path of $\frac{1}{2}\lambda$ by reflection at a glass surface.

For modifications of the Fresnel three-mirror system see Quincke[3] and Mascart.[4]

(ix.) *Lloyd's Mirror.*—This is a somewhat simpler arrangement[5] than the previous one. Interference takes place between the direct beam from a source S (*Fig.* 9), and the beam reflected from a single mirror (a piece of black glass is suitable) at almost grazing incidence, the second beam appearing to come from the virtual image S′. In this case again the central band of the interference system is

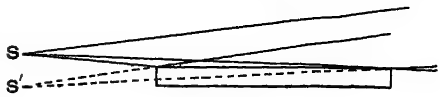


FIG. 9.

dark. Owing to the asymmetry of the optical system only one half of the fringe system can be seen, unless a thin transparent plate be held in the path of the direct beam so as to introduce retardation and consequent displacement of the fringes.

§ (5) ACHROMATIC FRINGES.—We have seen that the distance D between adjacent fringes is given by

$$D=\frac{x\lambda}{d},$$

from which it is clear that D varies with the wave-length λ. If it were possible to arrange that the ratio λ/d should be constant for all wave-lengths, one would be able to obtain a system of achromatic interference fringes when using white light, for then the fringes due to the different colours would coincide. A simple method of obtaining such achromatic fringes is to use as the two sources a short spectrum and its virtual image formed by reflection in a glass plate (Lloyd mirror system). The blue end of the spectrum should be towards the plate, as the distance between the sources for blue light should be less than for red light. If the spectrum is formed by a prism it is only possible to achromatise the fringes for two colours, but if a diffraction grating is employed the superposition of fringes for two wave-lengths will secure the superposition of fringes for all other wave-lengths.

§ (6) INTERFERENCE IN FILMS AND PLATES.—The interference phenomena which we have so far discussed can only be observed when a point source or a narrow line source of light is used. There is another group of interference phenomena which can be observed when using an extended source of light. In this group are included the cases of interference in transparent films and plates. A number of such cases will now be dealt with.

§ (7) INTERFERENCE IN THIN FILMS. (i.) *Simple Cases.*—When a very thin film of a transparent substance is illuminated by a broad source of light, such as a bright sky, brilliant colours are seen; soap bubbles and thin layers of oil spread over the surface of water are familiar examples of this phenomenon. These colours are due to interference between the rays reflected from the two surfaces of the film.

[1] G. Meslin, *Comptes Rendus*, 1893, cxvi. 250, 570; *Journ. de Physique*, 1893, (3), ii. 205.
[2] Complete circular fringes are not obtained, as the beams of light overlap only on one side of the optical axis, as may be seen from the diagram.
[3] G. Quincke, *Pogg. Ann.*, 1871, cxlii. 228.
[4] E. Mascart, *Comptes Rendus*, 1887, cv. 967.
[5] H. Lloyd, *Roy. Irish Acad. Trans.*, 1837, xvii. 174.

Let us first consider the case of monochromatic light falling on a thin film which is bounded by two parallel surfaces. At present we may confine our attention to the two paths ABCDE and FDE (*Fig.* 10). The

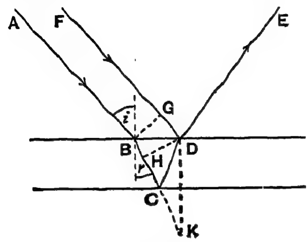


FIG. 10.

first is that of a ray which is refracted at the first surface, internally reflected at the second surface, and again refracted at the first surface; the second is that of a ray (emanating from the same point in the source) which is incident parallel to the first ray and is reflected at the first surface of the film. The difference in optical path of the two rays can easily be deduced. Let i be the angle of incidence and r the angle of refraction in the film. Draw BG perpendicular to FD and DH perpendicular to BC, and produce BC to meet the line through D normal to the film surface in K. Then BG and DH are the fronts of the incident and refracted waves respectively, so that the portions of the paths GD and BH are optically equal. Thus the path difference $d = \text{HC} + \text{CD} = \text{HK} = 2t\cos r$, where t is the thickness of the film. This is equivalent to $2\mu t\cos r$, measured in air, where μ is the refractive index of the film.

At first sight it would appear that there would be a maximum brightness when the path difference is equal to an even number of half wave-lengths and a minimum brightness when the difference is equal to an odd number of half wave-lengths This, however, does not fit in with the observed phenomena, and we can see that in the limiting case when the film is of zero thickness no light will be reflected, though we might from the above considerations have expected maximum brightness. The reason of this apparent discrepancy, as was first suggested by Young, is that light reflected at the surface of a denser medium suffers a phase change of 180°, which is equivalent to a path difference of half a wave-length. In the case of the film one of the rays is reflected without change of phase at the surface of a rarer medium, while the other is reflected at the surface of a denser medium. Thus the total retardation in the case under consideration is $2\mu t\cos r + \frac{1}{2}\lambda$. The brightness will therefore be greatest when $2\mu t\cos r$ is equal to an odd number of half wave-lengths, and least when $2\mu t\cos r$ is equal to an even number of half wave-lengths.

If now the incident light is white the light reflected from any point of the film will not include the wave-length which satisfies the equation $2\mu t\cos r = n\lambda$ at that point; the light will accordingly be coloured. If the film is of uniform thickness the colour will be uniform over the whole film.[1] Any want of uniformity in the thickness will give rise to variations in the colours observed at different parts of the film. This gives a convenient method of testing the uniformity of thickness of a film.[2]

(ii.) *Effect of Multiple Reflections* —The theory of thin films, as above sketched, is not complete, for we must take into account, as Poisson showed, the effect of multiple reflections. If d is the retardation of a ray once internally reflected, then $2d$ will be the retardation of a ray internally reflected three times, and so on. In order to sum up the effects of these multiple internal reflections it is necessary to investigate the relations between the amplitudes of the reflected rays. Let the amplitude of the incident ray AO (*Fig.* 11) be a, and the amplitudes of the reflected and refracted rays OB, OC be ab, ac respectively,

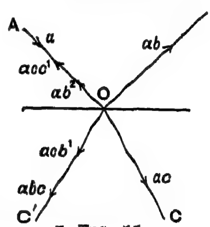


FIG. 11.

where b, c are fractions.[3] If these rays be reversed they should combine to give a ray along OA of amplitude a. The reversal of OC will give a ray of amplitude ab^2 along OA, and one of amplitude abc along the refracted path OC′. Let the components, along OA and OC′, of OC reversed be acc' and acb' respectively. Then

$$ab^2 + acc' = a \text{ and } abc + acb' = 0,$$

since there must be no resultant ray along OC′.

Thus

$$cc' = 1 - b^2 = 1 - b'^2 \text{ and } b = -b'.$$

We may calculate the resultant phase difference for the reflected beams, taking into account the multiple internal reflections. The amplitudes of the various beams are shown in *Fig.* 12. The common phase difference of the reflected components is given

[1] This is true only if the film is so thin that the variation of the angle of incidence from one point to another does not affect the value of $2\mu t\cos r$ by more than a small fraction. If a point source at infinity is used the statement is true for any thickness of film since the angle of incidence is then constant.

[2] See article on "Interferometers: Technical Applications."

[3] G. Stokes, *Math. and Phys. Papers*, ii. 89.

by $\delta=(2\pi/\lambda)d$, where d is the common path difference.

Now let A be the amplitude of the resultant reflected beam and Δ its phase when the phase of

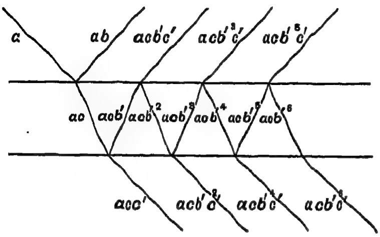


FIG. 12.

the first reflected beam is zero. Then, making use of the notation employed at the beginning of this article, we have

$$\mathrm{A}\cos\left\{\frac{2\pi}{\tau}\left(t-\frac{x}{\mathrm{V}}\right)+\Delta\right\}=ab\cos\frac{2\pi}{\tau}\left(t-\frac{x}{\mathrm{V}}\right)$$
$$+acb'c'\cos\left\{\frac{2\pi}{\tau}\left(t-\frac{x}{\mathrm{V}}\right)+\delta\right\}$$
$$+acb'^3c'\cos\left\{\frac{2\pi}{\tau}\left(t-\frac{x}{\mathrm{V}}\right)+2\delta\right\}+\ldots$$

Expanding, and equating the coefficients of

$$\cos\frac{2\pi}{\tau}\left(t-\frac{x}{\mathrm{V}}\right)\text{ and }\sin\frac{2\pi}{\tau}\left(t-\frac{x}{\mathrm{V}}\right),$$

we obtain

$$\mathrm{A}\cos\Delta=ab+acb'c'\cos\delta+acb'^3c'\cos 2\delta+\ldots,$$
$$\mathrm{A}\sin\Delta=\quad acb'c'\sin\delta+acb'^3c'\sin 2\delta+\ldots$$

Therefore

$$\mathrm{A}e^{i\Delta}=ab+acb'c'(e^{i\delta}+b'^2e^{2i\delta}+\ldots)$$
$$=ab+acb'c'\frac{e^{i\delta}}{1-b'^2e^{i\delta}}.$$

Multiplying the second term on the right above and below by $1-b'^2e^{-i\delta}$ we have

$$\mathrm{A}e^{i\Delta}=ab+\frac{acb'c'(\cos\delta-b'^2)}{1-2b'^2\cos\delta+b'^4}+i\frac{acb'c'\sin\delta}{1-2b'^2\cos\delta+b'^4}.$$

If we equate the real parts on both sides and the imaginary parts and square and add we get, on substituting

$$cc'=1-b^2\text{ and }b=-b',$$

$$\mathrm{A}^2=\frac{4a^2b^2\sin^2(\delta/2)}{1-2b^2\cos\delta+b^4}=\frac{4a^2b^2\sin^2(\delta/2)}{(1-b^2)^2+4b^2\sin^2(\delta/2)}.$$

By a similar calculation it is possible to deduce that the amplitude B of the resultant transmitted beam is given by

$$\mathrm{B}^2=\frac{a^2(1-b^2)^2}{(1-b^2)^2+4b^2\sin^2(\delta/2)}.$$

This result is to be expected, since A^2 and B^2 are the intensities of the reflected and transmitted beams, and their sum should equal a^2, the intensity of the incident beam.

The result for the reflected beam shows that the intensity is zero when

$$\delta=2n\pi,$$

that is, when $\quad 2\mu t\cos r=n\lambda,$

since $\quad \delta=\frac{2\pi}{\lambda}d=\frac{2\pi}{\lambda}\cdot 2\mu t\cos r.$

Thus, if light of wave-length λ is employed, the film or plate will appear perfectly black when $2\mu t\cos r$ is equal to an even number of half wave-lengths. This is the same result as was obtained when considering the simpler case of one internal reflection, except that there the condition was merely for minimum brightness.

It is interesting to note that the condition for zero intensity in the reflected beam corresponds to maximum intensity in the transmitted beam, and that the minimum intensity in the latter is not zero, but is such that when added to that of the reflected beam the sum is a^2. The phenomena in the two cases are complementary, but are much sharper in the reflected than in the transmitted beam.

(iii.) *Wedge-shaped Films.*—If the two plane surfaces bounding a film are inclined to each other at a small angle, a system of fringes will be observed. If white light is used the fringes will be coloured, but there is so much overlapping of the fringes that they will not be distinguishable unless the film is extremely thin and the angle of the wedge is very small. In the case, therefore, of films of appreciable thickness it is necessary to use monochromatic light, and the thicker the film the more monochromatic the light must be.

§ (8) NEWTON'S RINGS.—Newton observed that when a plane-glass surface was placed in contact with a convex spherical surface circular fringes were formed round the point of contact. These fringes, usually called Newton's rings, are due to interference in the film of air between the two surfaces. Now we have seen that where

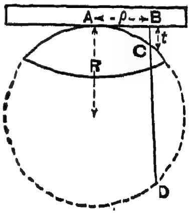


FIG. 13.

$$2\mu t\cos r=n\frac{\lambda}{2},$$

there will be maximum or minimum brightness according as n is odd or even, where μ and t are the refractive index and thickness of the film and r is the angle of refraction in the film. In the present case, if we consider the light to be incident normal to a plate of glass resting on the convex surface of a lens (*Fig.* 13), $\mu=1$ and $r=0$. Let A be the point of contact and R be the radius of

curvature of the convex surface. Then if t is the thickness of the air film at a point B, where $AB = \rho$, we have

$$\rho^2 = BC \cdot BD$$
$$= 2tR,$$

since we are considering points very close to the point of contact. Thus there will be a bright ring when

$$\rho = \sqrt{R(2n+1)\frac{\lambda}{2}},$$

and a dark ring when

$$\rho = \sqrt{Rn\lambda},$$

where n is any whole number. Newton's rings, therefore, give us the means of measuring R if λ is known.[1]

§ (9) Michelson's Interferometer.—A convenient method of observing interference phenomena in an air film of variable thickness was devised by Professor Michelson [2] The principle of the arrangement which he used is as follows. Light from a monochromatic source S (*Fig.* 14) falls on a plane-glass plate

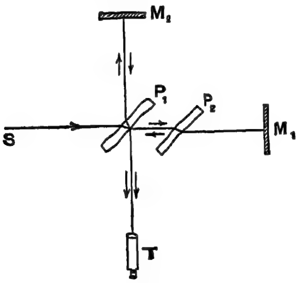


Fig. 14.

P_1 at an angle of 45°. The back surface of this plate is half silvered so that the beam of light is divided into two portions of approximately equal intensity. One of these, after transmission through P_1 and a second similar plate P_2 (unsilvered), and normal reflection at the plane silvered mirror M_1, passes again through P_2 and is reflected by the half-silvered surface of P_1 into the observing telescope T. The other portion, after internal reflection in P_1 and normal reflection at the plane silvered mirror M_2, is transmitted through the plate P_1 into the telescope. The reason for introducing the second plate, P_2, which is placed at 45° in the path of the first beam, is to make the glass paths of the two beams equal, for without it the first beam would pass once through P_1 whereas the second beam would pass through it three times. In order to obtain complete compensation of the glass paths the two plates, P_1, P_2, should be of equal thickness, preferably cut from the same piece of glass. If the air paths are now of the same length the arrangement is equivalent to a film of zero thickness, for the image of M_1 by reflection in P_1 will then coincide with M_2. One of the mirrors is provided with a micrometer screw by means of which it may be moved parallel to itself so as to introduce any desired difference in optical path. The plates and mirrors are also provided with fine adjustments to enable the image of M_1 to be brought into parallelism with M_2.

Modified forms of the Michelson interferometer are used for testing the planeness of surfaces, the homogeneity of glass blocks and prisms, and the aberrations in lenses.[3]

§ (10) Fabry and Perot Interferometer.—We have already seen, when considering the case of multiple reflections in films, that the intensities of the resultant reflected and transmitted beams are given by

$$A^2 = \frac{4a^2b^2 \sin^2(\delta/2)}{(1-b^2)^2 + 4a^2b^2 \sin^2(\delta/2)}$$

and $$B^2 = \frac{(1-b^2)^2}{(1-b^2)^2 + 4a^2b^2 \sin^2(\delta/2)}$$

respectively, where a^2 is the intensity of the incident light, b^2 is the reflection coefficient, and δ is the common phase difference between successive components of the reflected and transmitted beams. Confining our attention to the transmitted system, we see that, as δ varies, the intensity varies periodically; its maximum value, corresponding to $\delta = 2n\pi$, is a^2, its minimum value, corresponding to $\delta = (2n+1)\pi$ is $a^2((1-b^2)/(1+b^2))^2$, while an intermediate value, where $\delta = ((2n+1)/2)\pi$, is $a^2((1-b^2)^2/(1+b^4))$. It will be seen that, as the reflection coefficient increases, the minimum and intermediate values decrease, and, therefore, the maxima becomes sharper. *Fig.* 15 shows the types of intensity curves obtained,

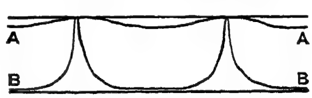


Fig. 15.

for normal incidence, with unsilvered (AA; $b^2 = \cdot 04$), and silvered (BB; $b^2 = \cdot 75$) plates. Thus by using a high reflection coefficient one obtains sharp bright lines on a dark background. The case for the reflected system is exactly reversed and is illustrated by *Fig.* 15 turned upside down. In this case,

[1] See article on "Spherometry."
[2] A. A. Michelson, *Am. J. Sci.*, 1882, xxiii. 395; *Phil. Mag.*, 1882, (5) xiii. 237.
[3] See article on "Interferometers: Technical Applications"; also F. Twyman, *Opt. Soc. Trans.*, 1920–21, xxii. No. 4.

if the reflection coefficient is high, one obtains sharp dark lines on a bright background.

The above fact has been made use of by Fabry and Perot[1] in their interferometer, which consists of two glass plates, the neighbouring faces of which are lightly silvered. One plate is fixed, and the second plate, which is set approximately parallel to the first, can be moved parallel to itself so that the thickness of the air film between the plates can be varied at will.

§ (11) Lummer and Gehrcke Interferometer.—The brightness of the reflected beams can also be increased by increasing the angle of incidence. Lummer and Gehrcke[2] have employed this principle in their application of interference methods to spectroscopy. They use a plate of glass (*Fig.* 16), at the end of one face of which is cemented a small right-angle prism. The incident light passes through the prism into the plate, and the beams of light which emerge, after successive internal reflections, from both sides of the plate can be examined by means of a telescope.

It might at first appear that the phenomena observable with the Lummer and Gehrcke

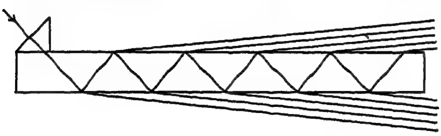


Fig. 16.

plate are due to diffraction effects, and this for two reasons. Firstly, there is a striking resemblance between the plate and the echelon grating, the path difference of successive beams increasing by definite steps. Secondly, it is well known that in interference systems such as we are considering the angular separation between the fringes decreases rapidly as the angle of incidence is increased, so that we might expect that in the case of the Lummer and Gehrcke plate, where the beams leave the plate at practically grazing emergence, the fringes could not be resolved. This, however, is not so, for, as one may easily calculate, the angular separation between successive fringes increases again as one approaches the limiting case of grazing emergence until it very nearly reaches the value for normal incidence. The phenomenon is, therefore, one of pure interference.[3]

§ (12) Interference in Multiple Thick Plates.—When two plates of glass are inclined to each other at a small angle (*Fig.* 17), a beam of light incident on one of the plates is divided into a number of components by multiple reflections between the plates. If the plates are of the same thickness some of these components are capable of interfering with each other. Thus in *Fig.* 17 the paths

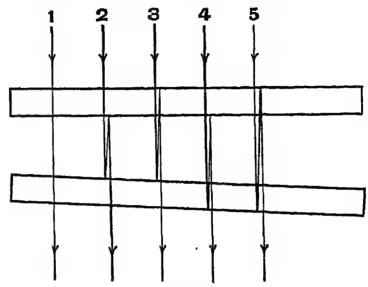


Fig. 17.

of the rays 3 and 4 are approximately equal, and, owing to the small inclination of the plates, such rays will form a series of interference fringes. Brewster[4] observed this phenomenon in 1817, and a modification of the arrangement has been employed by Jamin[5] in his interference refractometer. The principle of this instrument is illustrated in *Fig.* 18. A and B are two plates of optical glass, mounted parallel to each other and inclined at an angle of 45° to the line joining the centres of the surfaces a_1, b_1; the other surfaces, a_2, b_2, are silvered. If the plates have exactly the same thickness and the

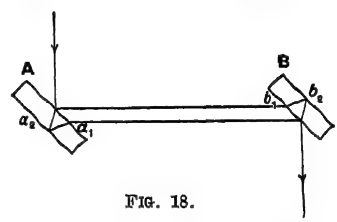


Fig. 18.

same refractive index, and are placed accurately parallel to each other, the optical paths of the two beams shown in *Fig.* 18 will be equal. On slightly rotating one of the plates about a horizontal or a vertical axis a path difference is introduced and a system of interference fringes can then be observed.

The Jamin refractometer can be used for measuring the refractive indices of gases or of transparent liquids or solids. Thus, for example, a plate of glass may be introduced into the path of one of the beams, in consequence of which there will be a displacement

[1] C. Fabry and A. Perot, *Ann. Chim. et Phys.*, 1897, xii. 459; 1899, xvi. 115; 1901, xxii. 546; *Comptes Rendus*, 1898, cxxvi. 331, 407, 1561, 1624, 1706, 1779.

[2] O. Lummer, *Verhandl. d. Deutsch. Phys. Ges.*, 1901, iii. 92; O. Lummer and E. Gehrcke, *Akad. Wiss. Berlin, Ber.*, 1902, 11; *Ann. d. Physik*, 1903, x. 457.

[3] Cf. R. W. Wood, *Physical Optics*, 1914, 283.

[4] D. Brewster, *Edin. Trans.*, 1817, vii. 435.

[5] J. Jamin, *Comptes Rendus*, 1856, xlii. 482; *Ann. Chim. et Phys.*, 1858, (3), lii. 163.

of the fringe system. The amount of this displacement depends on the thickness and refractive index of the plate, and may be determined by placing in the path of the other beam a compensator consisting of two exactly similar small-angle glass wedges, of known refractive index, which can be slid relatively to each other so as to form a plate of uniform but variable thickness. The compensator is adjusted until the centre of the fringe system is brought to its original position.[1]

In cases where it is desired to compare the refractive indices of similar gases, liquids, or plates of glass, that is, when the displacement of the fringe system is relatively small, one can employ a Jamin compensator for bringing the fringe system back to its zero position. This consists of two glass plates fixed to a common axis and inclined to each other at a small angle. If the compensator is set up so that each of the plates is in the path of one of the interfering beams, small relative retardations can be introduced by rotating the compensator about its axis. The amounts of these retardations may be calculated or may be obtained by calibrating the compensator by means of auxiliary experiments.

Modifications of the Jamin interference refractometer, necessitating the employment of only one glass plate, have been used by Mascart,[2] Lummer,[3] and Ketteler.[4] The system employed by Mascart and Lummer has recently been applied by Waetzmann[5] to the study of the aberrations in lens systems. Zehnder[6] and Mach[7] introduced two auxiliary mirrors, as illustrated in *Fig.* 19, in order to separate the interfering beams as much as possible.

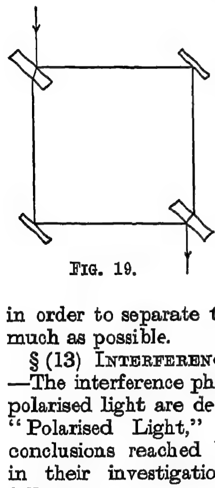


FIG. 19.

§ (13) INTERFERENCE OF POLARISED LIGHT.—The interference phenomena connected with polarised light are described in the article on "Polarised Light," *q.v.*, but the general conclusions reached by Fresnel and Arago[8] in their investigations may be stated as follows:

(i.) Two rays of light polarised at right angles do not interfere destructively under the same circumstances as two rays of ordinary light.

(ii.) Two rays of light polarised in the same plane interfere like two rays of ordinary light.

(iii.) Two rays polarised at right angles may be brought to the same plane of polarisation without thereby acquiring the quality of being able to interfere with each other.

(iv.) Two rays polarised at right angles, and afterwards brought to the same plane of polarisation, interfere like ordinary light if they originally belonged to the same beam of polarised light.

§ (14) APPLICATIONS OF INTERFERENCE PHENOMENA.—In addition to those mentioned in this article[9] the following applications of interference fringe systems to physical measurements, etc., may be mentioned: refractometry,[10] standardisation of wave-length and structure of spectral lines,[11] comparison of length gauges,[12] numerous physical and engineering problems,[13] and colour photography (Lippmann process).[14]

J. S. A.

LIGHT, INTERFERENCE OF, IN MULTIPLE THICK PLATES; JAMIN'S REFRACTOMETER. See "Light, Interference of," § (12).

LIGHT, PROPAGATION OF

HUYGHENS' PRINCIPLE

LIGHT is a form of energy which affects the nerves of the eye and is propagated by wave motion. The vibrations which constitute the waves are transverse to the directions in which they travel.

Since Fresnel introduced the conception of transverse wave motion a gradually increasing store of theoretical and practical evidence has been accumulated to demonstrate its truth. Until recently it was supposed that the mechanism concerned with the transmission of these vibrations resembled more or less that which regulates the propagation of transverse waves in an elastic solid. Recent discoveries have shown conclusively that light is an electromagnetic phenomenon, and that light waves obey the same laws as waves of electric and magnetic force. In the present article we are concerned chiefly with questions which are independent of the physical properties of the medium concerned in the transmission of the light. All we need to consider

[1] For the purpose of determining the centre of the system of monochromatic fringes a system of coloured fringes formed by white light is superimposed on the other system.
[2] E. Mascart, *Ann. Chim. et Phys.*, 1871, (4), xxiii. 146.
[3] O. Lummer, *Wied. Ann.*, 1884, xxiii. 513.
[4] E. Ketteler, *Pogg. Ann.*, 1871, cxliv. 372.
[5] E. Waetzmann, *Ann. d. Physik*, 1912, xxxix. 1042.
[6] L. Zehnder, *Zeits. Instrumentenk.*, 1891, xi. 275.
[7] L. Mach, *Zeits. Instrumentenk.*, 1892, xii. 89.
[8] A. Fresnel and F. Arago, *Ann. Chim. et Phys.*, 1819, (2), x. 288.

[9] See also article on "Interferometers: Technical Applications."
[10] Lord Rayleigh, *Phil. Mag.*, 1917, xxxiii. 161.
[11] E. C. C. Baly, *Spectroscopy*.
[12] F. Göpel, *Zeits. Instrumentenk.*, 1920, xl. 3.
[13] C. Barus, "The Interferometry of Reversed and Non-reversed Spectra," Parts 1-4, *Carnegie Inst. Washington Publications*, 1917.
[14] R. W. Wood, *Physical Optics*, 1914, 176.

are the rapid periodic changes of some vector quantity which, like other vectors, can be represented by a straight line in that it has length and direction.

While it is frequently convenient to visualise this vector as an actual displacement, it must be borne in mind that the changes which constitute light are electromagnetic in their nature; we cannot explain all the complex phenomena involved by the simple vibrations of ether particles.

Consider now a luminous source, O, *Fig.* 1. Light vibrations are emitted in all directions from this source, travelling outwards in waves with definite velocity. At one instant the disturbance will have reached a surface such as ABC, and all points of that surface will be in the same state of motion or, put in technical terms, in the same phase. Such a surface is known as a wave front. After another interval, the disturbance will have travelled on to A′B′C′, and this will have become the front of the wave. Light energy starting from O is, after a time, distributed over the surface ABC, and later, after another interval of time, over A′B′C′. If the medium in which the light is travelling is isotropic, that is, has identical properties in all directions, the speed of the light will be the same whatever be the direction of its motion, and it is easily seen that the surface ABC and A′B′C′ are both spheres with O as centre. If, however, the speed with which the light travels is different in different directions, as in a crystal, this is no longer the case; the wave front is no longer a sphere, but will take some other form depending on the properties of the medium.[1] Suppose, now, the source of light to be a long way off, the curvature of the portion ABC of the wave front with which we are dealing becomes less and less as the distance from the source is increased, and ultimately we arrive at a plane wave front, travelling always parallel to itself with a velocity which—in a crystalline substance—depends on the direction of its motion, but in an isotropic substance is a constant, independent of the direction.

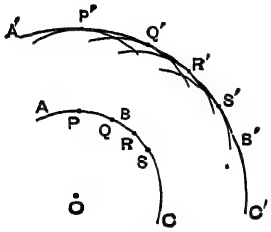


FIG. 1.

We may, however, regard the problem of the propagation of the waves from a rather different standpoint. At a given instant a certain state of disturbance exists over the surface ABC which, after a certain interval of time, has given rise to the disturbance existing over A′B′C′. How are these connected, and how can we pass from the conditions over ABC to those over A′B′C′? The answer to this question was given by Huyghens (Traité de la Lumière). Consider each point of ABC as the origin of a disturbance travelling outwards in a wave similar to that from the original source O, and draw the wave fronts corresponding to each of these points, P, Q, R, S, *Fig.* 1, and to the interval required by the principal wave to travel from ABC to A′B′C′. In the case of an isotropic medium these wave fronts will all be spheres; in a crystal they will have a more complicated form. In any case, however, it is possible to describe a surface which envelops or touches them all. According to Huyghens, this surface is the new position of the wave front; the disturbances due to the various secondary waves cancel each other everywhere, except on the envelope, while the resulting disturbance over the envelope is exactly that produced at the instant in question by the original wave-emanating front.

If we take, for simplicity, the case of a plane wave ABC in an isotropic medium, the secondary waves are clearly spheres of equal radii, with the various points of the wave as centre, while their envelope is a plane A′B′C′ parallel to ABC and at a distance from it equal to the radius of one of the secondary waves, that is, to the distance the disturbance has travelled in the interval of time under consideration.

This result is a consequence of the principle of Interference in Wave Motion, and though the arguments by which Huyghens supported it are not conclusive, it is shown in treatises on Theoretical Optics that the construction he gave is correct, and, when the form of the secondary waves is known, may always be applied to determine the form and position of the wave front, which represents the state of the primary disturbance.

We may also invert the procedure and apply it to finding the form of the wave surface diverging from a point in the following manner.

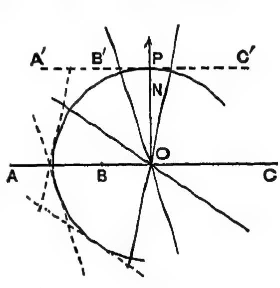


FIG. 2.

Consider a large number of plane waves all passing through the same point O, *Fig.* 2, but having various directions. After a given interval of time, t seconds say, each of these will have moved parallel to itself through a

[1] See "Light, Double Refraction of."

distance proportional to t and to the velocity of light measured in the direction in which the plane is moving. We shall have a series of planes corresponding with the original series but arranged round O—if the velocity of light in the medium is the same in all directions, these planes will all be equidistant from O—more generally they will be at varying distances, depending on the varying velocity of the light. In any case the planes will all envelop or touch a closed surface surrounding O, and this surface will determine the position which a disturbance emanating from O has reached in the time t seconds.

Moreover, if a wave front A′B′C′ corresponding to the wave ABC which passes through O touch the surface at P, the line OP constitutes the ray corresponding to this wave front and determines the direction in which a disturbance at O in the plane ABC is propagated, while ON, drawn at right angles to ABC, is the wave normal. If the medium is isotropic, the various planes such as A′B′C′ are all equidistant from O; the envelope of these planes, the wave front at time t, is clearly a sphere and the ray OP and the wave normal ON coincide.

In general this is not the case, and we have to distinguish between ON which determines the direction in which the wave travels, and OP the ray the direction in which the disturbance at O in the wave front ABC is propagated.[1] A distinction must also be drawn between the wave velocity, the velocity, that is, with which the wave front moves forward along ON, and the ray velocity, the velocity with which the disturbance travels along OP.

LIGHT, RECTILINEAR PROPAGATION OF

When the state of motion of an electrified particle is altered, a pulse of electric and magnetic force is emitted from the particle and travels out into space at a high velocity. This pulse carries with it a definite amount of electromagnetic energy.[2] If the motion of the particle be periodic the single pulse becomes a series of pulses. Energy is propagated outwards from the particle in waves; if the frequency in the waves lies within certain limits,[3] they affect the nerves of the eye and are known as light waves, producing the sensation of sight.

If we are dealing with a small luminous source O (*Fig.* 1) in an isotropic medium, the waves travel outwards in spheres from O. Let BAC be the position of a wave front at a time t. After an interval of time the light reaches an eye at E. The effect at E is due to the disturbance which at the previous time t existed over the surface BAC, and may be found by calculating the effect due to

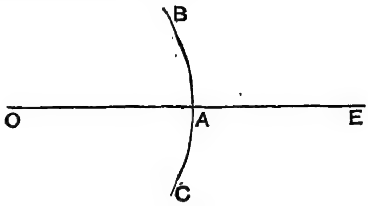


Fig. 1.

each element of that surface and taking the sum.[4]

But the eye does not recognise the surface BAC as the source of the light; it appears to come in a straight line from O, and the only portion of BAC which is effective in illuminating E is the portion A which lies directly between O and E. This rectilinear propagation is, it can be shown, a direct consequence of the fact that the wave-length of light is very small compared with the other distances concerned; it ranges from $\cdot 8\mu$ in the red[5] to $\cdot 4\mu$ in the violet. While the complete proof of this is a question of complicated mathematics, the following argument will indicate in a general way the reasons for it.

Imagine the source to be at a long distance behind the wave so that the portion of the spherical wave considered is practically plane. Let EA (*Fig.* 2), drawn perpendicular to the

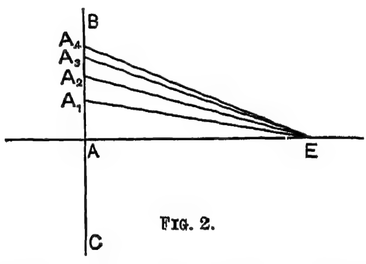


Fig. 2.

wide front BAC, be equal to a, and let λ be the wave-length. Take a series of points A_1, A_2, A_3 . . . on the waves at distances $a+\frac{1}{2}\lambda$, $a+\lambda$, $a+\frac{3}{2}\lambda$, . . . etc., from E, and describe a series of circles on the wave front with centre A and radii AA_1, AA_2, etc., thus dividing the front into a series of zones or rings. These are known as half-period elements. The disturbance which reaches E from A_1 is opposite in phase to that from A and therefore interferes with it; that from A_2 is in the same phase as that from A but opposite to that from A_1; thus, on the whole, the

[1] See "Light, Double Refraction of."
[2] See "Wireless Telegraphy," § (1), Vol. II.
[3] "Eye, The," § (6), etc.

[4] See "Light, Propagation of."
[5] The value of μ is 10^{-6} metres.

effect of the second half-period element is opposite to that of the first, while that of the third is of the same sign as the effect from the first. We may thus write the effect at E as

$$m_1 - m_2 + m_3 - \ . \ . \ ., \text{etc.}$$

The quantities m_1, m_2, etc., will be proportional to the areas of the consecutive half-period elements; they will also depend on the angles between the lines EA_1, EA_2, etc., and the direction of propagation, and will fall off slightly as these increase, while the increasing distance of each element from E will lead to the same result. Now to find the outer radius AA_n of any ring we have

$$EA_n^2 = AA_n^2 + b^2,$$

$$AN_n^2 = \left(b + \frac{n\lambda}{2}\right)^2 - b^2 = bn\lambda + \frac{n^2\lambda^2}{4}$$
$$= bn\lambda \text{ approximately,}$$

since λ^2 may be neglected.

Thus the area of the nth ring is given by $\pi(AA_n^2 - AA_{n-1}^2)$, and this is equal to $\pi b\lambda$, which is the same for all the rings.

Thus to our approximation, so far as the areas of the rings go, the quantities m_1, m_2, etc., representing the amplitudes of the disturbances at E due to the consecutive rings are equal; they decrease, however, in consequence of the increase of the angle AEA_n and of the distance EA_n as we get further from A. Thus in our series

$$S = m_1 - m_2 + m_3 \ . \ . \ . \ ;$$

the terms gradually decrease, but the difference between consecutive terms is small. The sum of such a series can be shown to be half the first term, for

$$S = \tfrac{1}{2}m_1 + \tfrac{1}{2}\{(m_1 - m_2) - (m_2 - m_3) + (m_3 - m_4) \ . \ . \ . \ \};$$

and the quantity within the bracket {. . .} is negligibly small, since each term is very small and the terms are alternately positive and negative.

Thus the amplitude of the disturbance at E due to the whole wave is just half that due to the first half-period element; the effective portion of the wave is limited to a small area about the point in which it is cut by the line joining the eye to the source of light. Light travels in a straight line from the source to the eye.

LIGHT, SOURCES OF, for microscopy with ultra-violet light. See "Microscopy with Ultra-violet Light," § (5).

Suitable for polarimetric work. See "Polarimetry," § (16).

LIGHT, SOURCES USEFUL FOR PROJECTORS, detailed consideration of. See "Projection Apparatus," § (5).

LIGHT FROM THE SKY. See "Scattering of Light by Gases," etc., § (1).

LIGHT SCATTERED BY DUST-FREE AIR, POLARISATION AND INTENSITY OF. See "Scattering of Light by Gases," etc., § (4).

LILIENFELD TUBE: an X-ray tube of the hot-cathode type developed in Germany. See "Radiology," § (14).

LIME, USE OF, IN GLASS MANUFACTURE. See "Glass," § (6).

LIPPICH'S POLARISER: the polarising system in use in many important forms of polarimeter at present. See "Polarimetry," § (6).

LISSAJOU'S FIGURES: a series of curves obtained by compounding simple harmonic vibrations used for testing the tuning of some simple interval between two forks. See "Sound," § (53) (v.).

LOG LINE, for measuring ships' speed. See "Navigation and Navigational Instruments," § (16).

LONGITUDE, DEFINITION OF. See "Surveying and Surveying Instruments," § (5).

LUMEN: the unit of luminous flux. See "Photometry and Illumination," § (2).

LUMETER: a portable illumination photometer. See "Photometry and Illumination," § (62).

LUMINESCENCE OF LUMINOUS COMPOUNDS, CAUSE OF. See "Luminous Compounds," § (7).

LUMINESCENT MATERIALS, USES OF. See "Luminous Compounds," § (8).

LUMINESCENT SUBSTANCES, light emitted by, when exposed to α-rays; particularly zinc sulphide. See "Luminous Compounds," § (2).

LUMINOUS COMPOUNDS

THE discovery of self-luminous compounds dates back to the days of alchemy. They were originally prepared at Bologna by calcining crystals of barium sulphate with organic matter such as flour. This operation converted the sulphate into the sulphide which after exposure to daylight appears luminous when placed in the dark. The salt was known as "Bolognian phosphorus" owing to its similarity as regards luminescence to phosphorus which was discovered before it. Another luminous compound was prepared by roasting oyster shell with sulphur. This yielded calcium sulphide and was known as "Canton's phosphorus." This salt was greatly improved by Balmain, who found that very minute quantities of other substances such as bismuth mixed with it added greatly to its luminosity. This substance emits a bluish light for several hours after exposure to light. It was first employed about 1877 on watch and clock dials, but was practically discarded about twenty years ago, largely

for the reason that it was necessary to expose the article to light to render the luminous paint effective.

§ (1) Radioactivity.—After the discovery of radium and other radioactive substances it was found that the rays from these substances were capable of exciting certain materials to the emission of visible light, even when those materials were not previously exposed to light. Becquerel[1] examined the luminescence produced in a number of materials when exposed to the rays from radium. The substance to be studied was placed in the form of powder above the radium on a thin sheet of mica. The following table shows the relative intensities of the luminescence excited in various substances:

Substance.	Intensity.	
	Without Screen.	With Screen of Black Paper.
Hexagonal zinc blende .	13·35	0·04
Barium platino-cyanide	1·99	0·05
Diamond	1·14	0·01
Double sulphate of uranium and potassium	1·00	0·31
Calcium fluoride . .	0·30	0·02

The values of the intensities given in the second column are those when the rays from the radium fall on the powder after passing through the thin sheet of mica supporting the powder. The mica sheet is so thin that only a small percentage of the α-rays emitted by the radium is absorbed. The intensity of the luminescence produced by the rays after passing through a sheet of black paper in addition to the mica is shown in the last column and is in each case given relative to that without the screen taken as unity. The sheet of paper is sufficiently thick to absorb most of the α-rays. It will be observed that there is a great diminution in luminosity when the rays are passed through the paper, which shows that the luminescence produced without the screen is mainly due to the α-rays. The material which shows the greatest effect is zinc sulphide—so-called "Sidot's hexagonal zinc blende." This gives by far the greatest luminescence of all materials hitherto examined when exposed to rays from radium.

§ (2) Luminescent Substances.—The light emitted by different materials when exposed to α-rays differs very markedly. Zinc sulphide preparations have a yellowish-green hue; willemite gives a very bright green hue; calcium tungstate has a bluish-green tint; cadmium phosphate a decided red colour. Calcium sulphide is very similar to willemite, but its phosphorescence is more persistent; calcium salicylate is similar to calcium tungstate.

[1] *Comptes Rendus*, 1899, cxxix. 912.

Some of these substances become luminous under the action of X-rays and ultra-violet light, but the effect due to these rays gives no definite guidance as to the effect to be expected when exposed to the rays from radium. For instance, the double sulphate of uranium and potassium is more luminous than hexagonal zinc sulphide under the action of X-rays, but the reverse is true of radium rays; under the influence of these rays calcium sulphide gives a blue luminescence, but is hardly affected by X-rays. Hence substances which are very luminescent under the action of X-rays do not necessarily produce the brightest luminous mixtures when radium is used.

The luminosity of zinc sulphide, and in fact that of all the other responsive salts, depends upon the state of its purity. Very pure zinc sulphide shows but little luminescence under the action of the rays from radium. It requires a trace of an impurity in it to produce the best effect. The luminosity depends upon the nature and the amount of this impurity—the colour of the luminescence also changes as the impurity is varied. There are therefore grades of zinc sulphide, depending upon the nature and amount of impurity present, which vary greatly in their response to the rays.

To test the suitability of any given material as a base for a luminous compound it is necessary either to add radium directly to the substance and compare the resulting luminosity with that of a good grade of zinc sulphide or else to bring the α-rays to play on the materials in a manner similar to that employed by Becquerel in his original investigation. The following method of examination is due to Viol and Kammer[2] and has been found very satisfactory. A plate of polonium about 20 millimetres in diameter is prepared which serves as an intense source of α-rays. With a suitable polonium plate a superficial luminescence can be obtained in zinc sulphide equal to that obtained by the addition of several hundred microgrammes of radium element per gramme of zinc sulphide. By holding this plate over two adjacent specimens of zinc sulphide or other material it is possible to determine with ease and rapidity the relative values of their α-ray luminescence.

At present the responsive base most commonly used for luminous compounds is zinc sulphide and the luminosities of other compounds are usually referred to that of a standard grade of this material.

§ (3) Preparation of Zinc Sulphide Luminous Compound.—There are two methods[3]

[2] *Am. Electrochem. Soc. Trans.*, 1917, xxxii. 381.

[3] The writer is indebted to Mr. F. H. Glew for details of these methods of preparation.

employed of mixing the radium salt with the sulphide, which may be called the dry method and the wet method respectively. Each of these methods has its own advantages, but of the two the latter is the more satisfactory.

(i.) *The Dry Method.*—Suppose it is desired to make a luminous compound containing a specified amount of radium per gramme of compound. A tube of known radium content is taken and its contents transferred into a well-polished agate mortar. The amount of zinc sulphide to be mixed with this quantity of radium to make a compound of the required strength is weighed out and placed in a watch-glass near at hand. When transferring the radium from the tube to the mortar it is difficult to remove all the salt, as some of it adheres to the walls of the tube even though the latter is tapped during the operation. Most of the remaining salt may, however, be removed by putting a little of the zinc sulphide into the tube and rubbing it on the inside walls by means of a small camel-hair brush. Only a very small quantity of the sulphide need be used for this purpose. The operation may be repeated five or six times so as to ensure that most of the material clinging to the walls is removed. Having now transferred the radium into the mortar it is ground for about twenty minutes into a fine powder. A small quantity of zinc sulphide is afterwards added, say about 1 gramme, and mixed with the radium by means of a spatula; no pressure is applied to the salt so as to preserve the zinc sulphide crystals, which should be kept as large as possible to obtain the best results. More sulphide is added and again thoroughly mixed with the mixture already in the mortar. If the quantity of zinc sulphide is not too large it may all be added little by little to the radium, stirring constantly and keeping about two grammes of sulphide in hand for final washings of camel-hair brush, spatula, and funnel used to transfer the compound from mortar to bottle. It is advisable not to fill the bottle too full so that the compound may be well shaken in it for final mixing.

(ii.) *The Wet Method.*—For this method it is necessary to use a radium salt which is soluble in water, the bromide or the chloride is suitable for the purpose. A tube containing a measured amount of one of these salts is taken and its contents dissolved in a known volume of distilled water with a trace of hydrochloric acid added to ensure that all the salt passes into solution. Suppose it is necessary to make a certain amount of luminous compound of a given strength. A volume of the solution containing the required amount of radium is measured out and placed in a shallow porcelain dish. If the volume of solution is greater than that necessary just to moisten the whole mass of the zinc sulphide required to make up the compound, it is advisable to evaporate off some of the water. Having reduced the volume of the solution the whole of the zinc sulphide is slowly added to it until all the liquid is absorbed. The dish is now placed in a water bath and allowed to remain there until the salt is thoroughly dry, occasionally lightly stirring the whole mass with a glass rod to break up any lumps that may be formed. When quite dry any small lumps may be smoothed out on a piece of ground glass by means of a flexible palette knife of gold, platinum, or bone, afterwards mixing thoroughly on the ground glass. Very little pressure should be used during this process, in order to avoid breaking up the zinc sulphide crystals. This method ensures that the radium is uniformly mixed with the sulphide and in this respect is superior to the dry method.

The dry method of preparation is very useful in the case of salts that are insoluble or only slightly soluble in water, but the wet method is more accurate. Also, whereas a given quantity of luminous compound of any desired strength can be readily made by the latter method, the quantity of compound of given strength which can be made by the former method depends upon the amount of radium which the tube to hand contains.

§ (4) Application of the Compound to Dials, etc.—The radium compound can be applied to dials, etc., by mixing the powder with some binding material such as celluloid, crystal, or copal varnish. The best method of mixing is to place a little heap of the compound in the form of powder in the middle of a watch-glass, and then to add a few drops of turpentine to the base of the heap. The turpentine is drawn up by capillarity into the body of the powder driving the air out in front of it. Varnish is then added, the minimum possible quantity being used, and the whole thoroughly mixed with the brush which is finally going to be used for the application of the paint. It is most readily applied by means of a fine sable brush. In painting a dial, for instance, successive coats of the paint are applied, each coat being allowed to dry thoroughly before putting on the next. It will be seen later that the addition of the varnish cuts down the luminosity of the powder by about 75 per cent even though the minimum amount of varnish be employed in the mixing. It is important, therefore, to keep the amount of varnish as small as possible. If the object painted is not protected by a glass cover or by some other means, it is advisable to give it a final coat of varnish for a protection. When the paint is to be applied directly to silver, iron, brass, or copper, a preliminary

coating of zinc white mixed with the varnish should be given. It is inadvisable to use white lead for this purpose because zinc sulphide reacts with this compound producing a dark sulphide which will reduce the luminosity of the paint.

Painted surfaces should not be submitted to measurement of luminosity for at least seven days after the final coat of varnish has completely dried. The manipulation of the compound in powder and liquid form permits the escape of emanation, the presence of which enhances the luminous effect; after a time this again accumulates, but it does not reach a state of radioactive equilibrium on account of the porosity of the varnish. A sufficient amount, however, is held by the varnish to give a very appreciable increase in the luminosity after a few days' time.

§ (5) Measurement of Radium Content.—The radium content of a luminous compound is usually measured by one of the γ-ray methods (see under "Radium"). The compound, after being made, is hermetically sealed in a glass tube and put aside for about four weeks before it is compared with the radium standard. During this interval emanation is accumulating and this reaches a state of radioactive equilibrium with its products in about a month's time. The curve shown in *Fig.* 1 shows the

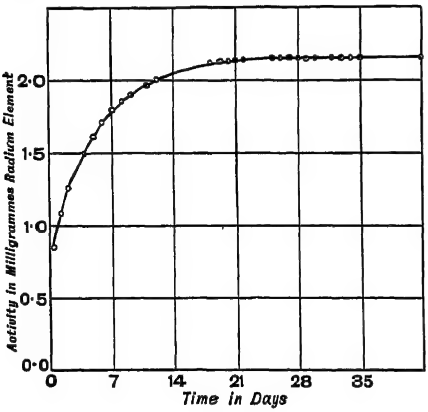


Fig. 1.

increase of the γ-ray activity of a tube of radium compound prepared by the dry method from the day it was sealed up until it reached a state of radioactive equilibrium. The amount of radium per gramme of compound is generally small and it requires a few grammes of the compound in order to carry out an accurate comparison with the standard. On this account the γ-rays emitted by the radium in the compound have to pass through a considerable thickness of material before they leave the tube, and the absorption taking place has to be corrected for in deducing the true radium content from the actual observations. Suppose that the tube containing the radium compound is cylindrical. Let I be the measured intensity of the γ-radiation emitted and I_0 the intensity if no absorption had taken place. Neglecting the correction due to the obliquity of the rays it can be readily shown that [1]

$$\frac{I}{I_0}=\frac{1}{\mu a\sqrt{\pi}}\sum_{r=1}^{r=\infty}(-1)^{r+1}\frac{(2\mu a)^r}{\underline{|r+1}}\cdot r\cdot\frac{\Gamma r/2}{\Gamma(r+1)/2}$$

$$=C_0+C_1(\mu a)+C_2(\mu a)^2+\ldots,$$

where μ is the coefficient of absorption of the rays per centimetre thickness of the compound and a the radius of the tube in cm.

The values of the constants C_0, C_1, C_2 . . . are given in the following table:

Coefficient.	Value.	Coefficient.	Value.
C_0	+1·000000	C_7	−0·001642
C_1	−0·848826	C_8	+0·000347
C_2	+0·500000	C_9	−0·000067
C_3	−0·226354	C_{10}	+0·000012
C_4	+0·083333	C_{11}	−0·000002
C_5	−0·025869	C_{12}	+0·000000
C_6	+0·006944		

The mean value of the absorption coefficient of the rays in zinc sulphide in the powdered form may be taken as 0·10 per cm. Also the absorption coefficient of the rays in soda glass is 0·10 per cm., so that we have the following table giving the ratio of the true to the apparent or measured radium content for tubes of different external diameters:

External Diameter of Containing Tube. Cm.	$\frac{\text{True Content.}}{\text{Apparent Content}}$
0·5	1·02
1·0	1·04
1·5	1·065
2·0	1·09
2·5	1·11
3·0	1·13

§ (6) Luminosity.—The luminosity of radium compound is measured by the photometric method.[2] The luminosity falls off with time. This loss of brightness is not due to any change in the radium, because it takes about 1700 years for the radium to decay to half value. The cause is to be sought for in the deterioration of the zinc sulphide, brought about by the continued bombardment of the crystals by the α-particles.

(i.) *Rate of Decay of Luminosity of Dry Material.*—An extensive investigation of the

[1] See Owen and Fage, *Phys. Soc. Proc.*, 1921, xxxiv. I. 27.
[2] See "Photometry and Illumination," § 125.

loss of brilliancy of radium compound has been carried out by Paterson, Walsh, and Higgins.[1] *Fig.* 2 is a curve which shows the variation with time of the luminosity of a sample of compound containing 0·2 mgm. radium element per gramme. The same type of curve is obtained whether the sample is prepared by the dry or by the wet method. After attaining its maximum value at the end of about 10 to 20 days after mixing the luminosity begins to decrease, at first very gradually and then more rapidly, until, after a period of about six to seven weeks from the date of mixing, the rate of decay appears to obey an exponential law, that is, the ratio of the rate of decay of luminosity at any time to the luminosity at that time is constant, so that if B is the luminosity at any time t, and B_0 the initial luminosity, then $B = B_0 e^{-kt}$, where k is a constant. At the end of about 200 days, however, the luminosity begins to diminish at a slower rate than that to be expected from the above law. This

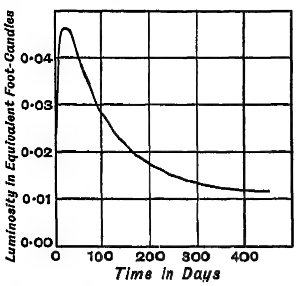


FIG. 2.

would indicate that there is an effect superposed on the normal decay operating in the reverse direction, which becomes relatively more and more important as the luminosity becomes smaller. After about 500 days the luminosity attains a value which is sensibly constant. This period, however, is not the same for all compounds, as in some the rate of decay diminishes more rapidly than in others.

(ii.) *Effect of Radium Content.*—The effect of the amount of radium present per gramme of compound on the rate of decay of luminosity is shown in *Fig.* 3. The curves refer to two samples of compound prepared at the same time with an identical quality of zinc sulphide, the one sample (top curve in figure) containing twice as much radium per gramme as the other. The initial brightness is approximately in the direct ratio of the radium content, but the rate of decay of compound of smaller content is slower, so that the ratio of the luminosities of the two samples gradually approaches unity as time goes on. Hence if a compound is to be used for a long period of time, there is very little advantage to be gained, except from the point of view of initial brightness, by increasing the radium content beyond a certain value. For short-period use, say three to four months, the radium content may

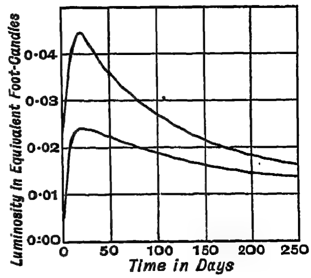


FIG. 3.

be usefully as high as 0·2 milligramme of radium element per gramme of compound. If the compound is likely to be required for more than six months, the use of a compound containing more than 0·1 milligramme of radium element per gramme of compound can only be justified by the need of greater brilliancy during the early part of the life.

(iii.) *Effect of Thickness.*—The effect of thickness of material on the luminosity is shown[2] in *Fig.* 4. The curves refer to a compound containing 0·15 milligramme radium element per gramme. When freshly made the ratio of the luminosity of a thickness of 0·3 mm. of compound to that of a thickness of 1·5 mm. is 0·63 (curve A). After an interval of 100 days (curve B) this ratio increases

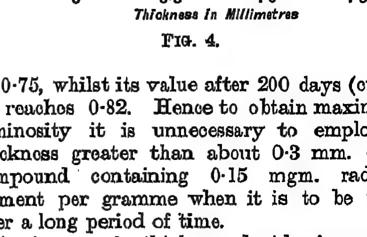


FIG. 4.

to 0·75, whilst its value after 200 days (curve C) reaches 0·82. Hence to obtain maximum luminosity it is unnecessary to employ a thickness greater than about 0·3 mm. of a compound containing 0·15 mgm. radium element per gramme when it is to be used over a long period of time.

An increase in thickness, besides increasing up to a certain point the initial brightness, also tends to accelerate the decay of luminosity—the effect is similar to that of increasing the

[1] *Phys. Soc. Proc.*, 1917, xxix. 215.

[2] Clinton, *Illum. Eng.*, 1918, xi. 260.

percentage amount of radium in the composition. This can be explained by the fact that in the greater thicknesses the front layers of material will receive additional bombardment from the radium in the deeper layers, and will therefore glow brighter, but have a shorter life, than if a thinner layer of material were used.

(iv.) *Rate of Decay of Luminosity of Varnish Material.*—The above results apply to luminous compounds in the dry state. When mixed with varnish and applied to dials, etc., the rate of decay of luminosity is different. The results obtained with six metal dials used for aeroplane instruments are given in the following table:

Dial.	Initial Luminosity (per cent of Standard).	Luminosity after Twelve Months (per cent of Standard).	Decay of Luminosity (per cent of Initial Luminosity).
1	94	59	63
2	82	45	55
3	94	57	61
4	63	33	52
5	91	39	43
6	52	24	46
		Average . .	53

The dials were all ordinary metal dials painted with a matt black surface, and having the luminous compound laid on in fine lines about 0·2 mm. to 0·3 mm. high to mark the graduations and figures of the dial. They were painted with different samples of luminous compound containing 0·2 mgm. radium element per gramme.

The rate of decay of luminosity of the dials is noticeably lower than that of the compound before application, the half-value period of the painted dial being about twelve months, whereas that of the dry powder is only about five months (see *Fig.* 2). The initial luminosity of the painted dial is also smaller, being of the order of $\frac{1}{3}$ or $\frac{1}{4}$ of that of the compound before application. The radium content of the varnish material may therefore with advantage be higher than that of the dry material. For short-period use it may be advantageous for some purposes to use varnish material containing as much as 0·3 mgm. or 0·4 mgm. radium element per gramme, but for long-period use the most suitable radium content is 0·2 mgm. per gramme, and maximum luminosity may be obtained by using one gramme of compound to paint an area of about 30 square centimetres.[1] It is important that the interval between the preparation of the compound in the dry state and its application to the dial should be as short as possible, as the deterioration in luminosity takes place from twice to three times as rapidly in the powder as in the painted dial. These differences between the dry and varnish compounds are due to the fact that the varnish impedes the activity of the α-ray bombardment from the radium and thereby causes an effective lowering of the radium content of the compound with a subsequent reduction of both the initial luminosity and also the rate of decay.

§ (7) Cause of Luminescence of Luminous Compounds.—Marsden[2] studied the effect of continued bombardment of α-particles on a screen of zinc sulphide and found that the actual number of scintillations produced by a constant source of α-particles changes very little with time of exposure, but that the brightness of the scintillations rapidly diminishes. Pure zinc sulphide does not exhibit the scintillation effect, but a salt containing an amount of impurity of the order of 1 per cent does. Assuming then that the scintillations are due to the presence of this impurity, only a small fraction of the total number of molecules are concerned in the effect. Rutherford[3] explained the scintillation effect by assuming that the impurity causes the presence of a number of "active centres" in the salt which are on the average uniformly distributed among the inactive molecules. When an α-particle strikes an active centre the latter is dissociated and emits light, but it is no longer effective in producing light when struck again by an α-particle. The effect of the continued bombardment by α-particles is thus to destroy gradually the active centres. Since probably many thousands of these centres initially lie in the path of the α-particle the effect of a continued bombardment will not cause an appreciable diminution in the number of scintillations, but it will diminish their intensities, for the intensity of a scintillation on the average will be proportional to the number of active centres remaining. These conclusions agree with the experimental results of Marsden. The theory in its simple form fails, however, to represent exactly all the facts which later experiments have brought to light. For instance, Paterson, Walsh, and Higgins (*loc. cit.*) found that the luminosity of a luminous compound at first rises rapidly to a maximum and then decreases according to an exponential law for about 200 days. After that time the rate of decrease of luminosity becomes slower and slower, the brightness tending to approach a limiting value which is not zero. Rutherford's theory gives a good approximation for the early period of decay, but fails increasingly at the later period. The theory has been extended by Walsh,[4] who assumes that the active centres which have been struck by α-particles, and according

[1] See Glew, *Opt. Soc. Trans.*, 1916, xvi. 276.

[2] *Roy. Soc. Proc.* A, 1910, lxxxiv. 548.

[3] *Ibid.*, 1910, lxxxiv. 561.

[4] *Ibid.*, 1917, xciii. 550.

to Rutherford's theory are put completely out of action from the point of view of luminescence, commence to recover according to an exponential law. With this modification the theory represents the observed facts much more accurately.

§ (8) Uses of Luminescent Materials.—Daylight luminous paints can only be used in exposed situations so that daylight can freely be absorbed to be given out again at night. Some use has been made of this class of luminous paint in shipping—luminous life belts are employed which may be clearly seen at night when thrown into the water; floating buoys can also be distinguished at night.

Much more use can be made, however, of luminous paints which are independent of daylight. Paper, porcelain and metals, dials of all sorts bearing figures and lines, are now used. In the case of some instruments, such as those used in aeroplanes, the requirements of luminosity are most important and outweigh other considerations as to cost and length of effective life. For such purposes grades of luminous paint are prepared which give the highest initial brightness. These have a high radium content and consequently a short effective life. For many other purposes such as that of illuminating marching compasses, dials of wrist watches, etc., a paint of lower initial luminosity is prepared, and this has the advantage of a lower cost and a longer life. For watches, push buttons, etc., of the better grade an effective life of ten years or more is desirable, and assuming that a luminosity of 25 per cent of the original is still satisfactory, a luminous paint containing about 100 microgrammes of radium element per gramme of compound should answer the requirement. Compounds with more radium than this would have a shorter effective life, and compounds with a smaller proportion of radium a longer effective life.

During the war much use was made of radium luminous compound in the dry powder form for gun-sights, to facilitate accurate night-firing. For this purpose the powder was sealed up in small flat glass tubes, their internal bore from back to front not exceeding about 0·4 mm., their other dimensions depending upon the class of instrument to which they were attached. For illuminating at night the cross wires of telescopic gun-sights attached to big guns, the powder was placed in glass tubes of about 0·5 mm. bore bent into circular forms to fit the metal tube of the gun-sight and situated near the plane of the cross wires. This device did away with the necessity of having glow-lamps for illuminating purposes with the attendant risk of detection by the enemy.

E. A. O.

Luminous Paint (Radium), Photometry of. See "Photometry and Illumination," § (125).

Luminous Power (or Intensity): a source of light. See "Photometry and Illumination," § (2).

Lummer-Brodhun Photometer: one of the most sensitive types of photometer head. See "Photometry and Illumination," § (18).

Lummer-Brodhun Spectrophotometer. See "Spectrophotometry," § (12).

Lummer-Kurlbaum Standard: an incandescence standard of light. See "Photometry and Illumination," § (10).

Lux: the unit of illumination; another name for the metre-candle. See "Photometry and Illumination," § (2).

Luxometer: a portable illumination photometer. See "Photometry and Illumination," § (61).

— M —

M Series: a group of spectrum lines in the characteristic X-rays emitted by an element. See "Radiology," § (17).

Maddox Set of Rods or Grooves: an effective apparatus for distorting one of the two images of an object seen by a subject's two eyes, so that the brain no longer attempts to associate it with the other image; this is the first step in the use of a phorometer. See "Ophthalmic Optical Apparatus," § (3).

Magnesia, Use of, in Glass Manufacture. See "Glass," § (9).

Magnetic Compass, description. See "Navigation and Navigational Instruments," § (9).

Mathematical theory. See *ibid.* § (10).

MAGNETIC ROTATORY POWER

§ (1) Faraday's Experiments.—The phenomenon of *magnetic rotation* of the plane of polarisation of light was discovered by Faraday in 1845, and described by him in a paper "On the Magnetisation of Light."[1] A ray of light issuing from an Argand lamp was polarised in a horizontal plane by reflection from a surface of glass, and the polarised ray passed through a Nicol's eyepiece revolving on a horizontal axis. Between the polarising mirror and the eyepiece two powerful electromagnetic poles were arranged; they were separated from each other by about 2 inches in the direction of the line of the ray, which passed either by the side of them or between

[1] *Phil. Trans.*, 1846, p. 1.

them in such a way that the ray was nearly parallel to the magnetic lines of force. A piece of dense optical glass (borosilicate of lead), which he had described in 1830, about 2 inches square and 0·5 inch thick, was placed lengthways between the poles. The eyepiece was turned to extinguish the polarised ray, but this was transmitted again when the current was passed through the electromagnet. The image of the lamp flame was, however, extinguished again by rotating the eyepiece to the right or to the left according to the direction of the current in the electromagnet.

The rotation was found to be proportional to the intensity of the magnetic force and to the length of the column of glass exposed to it. Smaller rotations were produced by other glasses, by some crystals, by water, alcohol, ether, various oils, and a large range of aqueous solutions, but no action was detected in the case of gases. When the medium possessed a natural rotatory power, the magnetic rotations were superposed upon (*i.e.* added to or subtracted from) the natural rotations.

§ (2) Apparatus and Methods of Measurement. — The type of apparatus used for

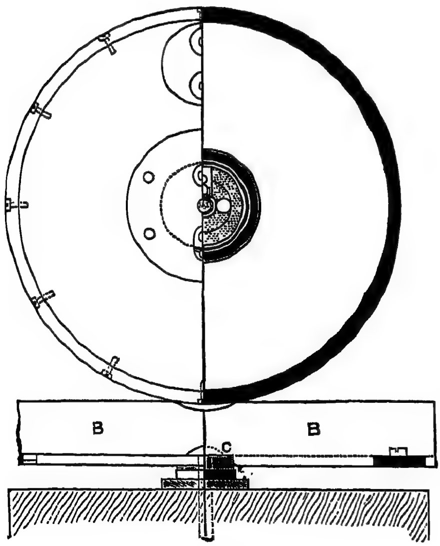


Fig. 1.

measurements of magnetic rotatory power is shown in *Figs.* 1 and 2, whilst the electrical connections are shown in *Fig.* 3.[1]

[1] *Trans. Chem. Soc.*, 1913, ciii. 1322 and 1331.

A coil of insulated copper wire A is wound in the form of a bobbin on a brass tube provided with a thin lining of cold water to

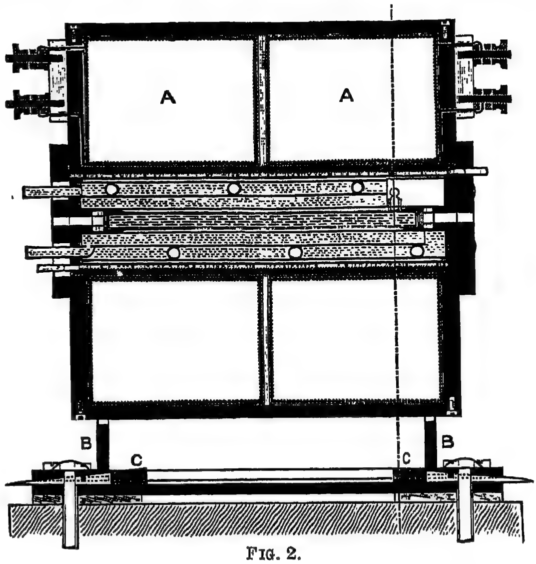


Fig. 2.

protect the material under examination from the heat of the magnet. The coil is enclosed by iron (or mild steel) end-pieces and a length of iron pipe; and movable pole-pieces are provided to lead the lines of force to the axis of the magnet. The magnet is mounted on rails, BB, so that it can be rolled away from the axis of the polarimeter, and wedges CC are provided for bringing the axis of the magnet into line with the axis of the polarimeter.

The substance to be examined (usually a liquid) is enclosed in a double water-jacket, connected to a thermostat. A correction must be applied for the discs at the end of the tube, since these produce a magnetic rotation of light; this correction is usually determined by introducing half-a-dozen discs together, and calculating the effect for a single disc; a group of discs, or a short rod of glass, may also be used to explore the strength of the magnetic field along the axis. In view of the variations which are usually found, it is convenient to standardise the apparatus by making measurements of the magnetic rotation produced by a tube of water, and using this as a basis of comparison for other liquids.

As the resistance of the magnet increases with rising temperature, a series of resistances is provided (*Fig.* 3) which can be cut out as the magnet becomes hot. Casual variations in the voltage on the mains are compensated by a sliding resistance and a carbon resistance, the constancy of the current being checked by

a null method, as shown in *Fig.* 3. Measurements are made with the current in each direction, the normal value of the readings

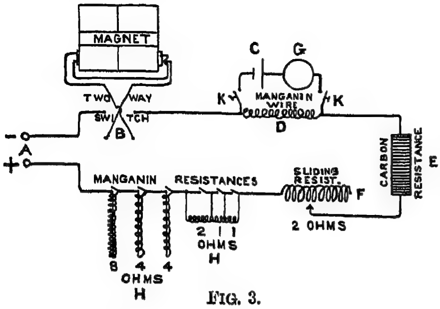


FIG. 3.

in the apparatus shown being about ±5°, or 10° in all.

In calculating the *molecular magnetic rotation* of a liquid, the formula generally used is

$$\text{Mol. rot.} = \frac{r \times M}{d} \div \frac{r_0 \times M_0}{d_0},$$

where r and r_0 are the observed rotations for the substance and for water, M and M_0 are their molecular weights, and d, d_0 their densities.[1]

§ (3) INFLUENCE OF THE MEDIUM.—The molecular rotations of a large number of compounds have been investigated, especially by Sir William Perkin.[2] The main result of these researches, which need not be described in detail, is to show that the molecular magnetic rotatory power of a substance is essentially an *additive property*, made up of the sum of the *atomic rotations* of the constituent atoms. This property is therefore closely analogous to the *molecular refraction* of a compound, but differs from it in that the influence of chemical constitution is more pronounced. Measurements of magnetic rotatory power can, therefore, be used to throw light on the constitution of compounds of doubtful molecular structure, *e.g.* in cases of dynamic isomerism, where ordinary chemical methods are not available.

§ (4) RELATION BETWEEN REFRACTION AND MAGNETIC ROTATION.—Drude[3] has deduced the following relationship. If the refractive index n can be represented with sufficient accuracy by the equation

$$n^2 = a + \frac{b}{\lambda^2 - \lambda_1{}^2},$$

then the magnetic rotation can be represented by the formula

$$\delta = \frac{1}{n}\left(\frac{a'}{\lambda^2} + \frac{b'\lambda^2}{(\lambda^2 - \lambda_1{}^2)^2}\right).$$

This formula was tested with satisfactory results in the case of carbon disulphide and of creosote (*sic*), but in view of the slender character of the data used for the test, and the fact that two arbitrary constants, a' and b', are available for adjustment, in addition to the three constants a, b, and λ, used to express the refractive power, the value of this formula does not appear to be very great.

§ (5) NATURAL AND MAGNETIC ROTATORY DISPERSION.—Six years after Faraday had discovered the "magnetisation of light," Wiedemann[4] made comparative measurements of the natural and magnetic rotations in turpentine for five wave-lengths in the solar spectrum, and concluded that the two rotations were proportional throughout the spectrum. This relation, which is generally known as *Wiedemann's Law*, has been verified in the case of quartz[5] and of sodium chlorate,[6] and is perhaps generally true for optically active crystals. In the case of optically active liquids, however, it generally fails, and the approximate agreement between the magnitude of the two dispersions, which is occasionally observed, is probably only a coincidence.[7] It is, however, important to notice that the simple dispersion formula

$$\alpha = \frac{k}{(\lambda^2 - \lambda_0{}^2)},$$

which serves to express the natural rotatory dispersion in a wide range of compounds, can be used equally well for magnetic rotatory dispersions, although its validity cannot be subjected to the same drastic tests, on account of the small magnitude of the readings which are usually obtained. Until more exact measurements are available, this simple two-constant formula may be used to represent the data for magnetic rotatory dispersion, in preference to Drude's formula, which involves five constants and demands a knowledge also of the refractive dispersion of the compound.

T. M. L.

MAGNIFICATION METHODS OF DETERMINING FOCAL LENGTHS. See "Objectives, Testing of Compound," § (2) (iii.).

MANDOLIN: a musical instrument of four double strings which are plucked with a plectrum and stopped with the left hand upon the fingerboard. See "Sound," § (27).

[1] Perkin, *Trans. Chem. Soc.*, 1896, p. 1060.

[2] *Trans. Chem. Soc.*, 1884, p. 421; 1886, pp. 205, 317, and 777; 1887, pp. 362 and 808; 1888, pp. 561 and 695; 1889, p. 680; 1891, p. 981; 1892, p. 800; 1893, pp. 57 and 488; 1894, pp. 20, 402, and 815; 1895, p. 255; 1896, p. 1025; 1902, pp. 177 and 292; 1905, p. 1491; 1906, pp. 33, 608, and 849; 1907, p. 806.

[3] *Theory of Optics*, 1907, ch. vii. p. 439.

[4] *Pogg. Ann. Phys. Chem.*, 1851, li. 82, 231.

[5] Lowry, *Phil. Trans.*, 1912, A, ccxii. 295.

[6] J. Dahlen, *Zeitschr. wiss. Phot.*, 1915, xiv. 315.

[7] Lowry, Pickard, and Kenyon, *Trans. Chem. Soc.*, 1914, cv. 95.

MANGIN MIRROR, for headlights. See "Projection Apparatus," § (10).
For searchlights. See *ibid.* § (12).
For signalling lamps. See *ibid.* § (11).

MAP PROJECTIONS. See "Surveying and Surveying Instruments," § (8).

MARTENS ILLUMINATION PHOTOMETER. See "Photometry and Illumination," § (60).

MARTENS POLARISATION PHOTOMETER. See "Photometry and Illumination," § (30).

MARTENS (KÖNIG-) SPECTROPHOTOMETER. See "Spectrophotometry," § (12).

MASS, ABSORPTION COEFFICIENT: a constant in X-ray measurements obtained by dividing the absorption-coefficient of an absorbing material by its density. See "Radiology," § (17).

MATTHEWS-DYKE PHOTOMETER: a photometer for the determination of average candle-power. See "Photometry and Illumination," § (44).

MATT SURFACE: one in which specular reflection is absent or very slight. See "Photometry and Illumination," § (51).

MAXWELL'S COLOUR-BOX: an apparatus for determining the data required to specify colours in terms of three primary colours. See "Eye," § (10)

MEAN HORIZONTAL CANDLE-POWER. See "Photometry and Illumination," §§ (2) and (38).

MEAN SPHERICAL CANDLE-POWER. See "Photometry and Illumination," § (2) and § (42) *et seq.*

MEAN-TONE TEMPERAMENT: a particular musical temperament which uses one size of tone only, but makes it the mean of the large and small tones required by the just intonation. See "Sound," § (6) (ii.).

MELTING OF GLASS, PROCESS OF. See "Glass," § (15).

MESOTHORIUM, methods of detection of presence of, in radium compounds. See "Radium," § (9) (iv.).

METRE-CANDLE: the metric unit of illumination, being the amount of light falling on a square centimetre placed in a direction normal to the light at a distance of one metre from a source of one candle-power. See "Photometry and Illumination," § (2).

MICA, EFFECT OF α-RAYS ON. See "Radioactivity," § (15).

MICHELSON'S absolute measurement of a light-wave by the interferometer method; by determining the number of waves of three radiations from cadmium vapour which were equal in length to the standard metre. See "Wave-lengths, the Measurement of," § (3).

MICHELSON'S INTERFEROMETER. Twyman's modification for testing optical instruments. See "Interferometer, Technical Applications," § (5).

MICROMETER MICROSCOPES. See "Divided Circles," § (13).

MICRO-POLARIMETER: a polarimeter designed to measure the optical rotation of substances that can only be obtained in small quantities. See "Polarimetry," § (15) (i.).

MICROSCOPE. Applications to curvature measurements. See "Spherometry," § (8).

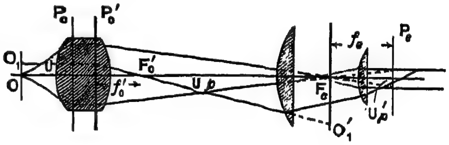
MICROSCOPE, OPTICS OF THE

I. INTRODUCTION

THE compound microscope consists of an objective which forms an inverted real image of the object under observation, and of an eyepiece by which the primary image is further magnified and either presented to the eye of the observer as an inverted virtual image, or projected upon a screen or photographic plate as an erect real image.

As the instrument can be usefully employed only within that part of its field for which the aberrations are sufficiently well corrected, the effects produced may be legitimately discussed with the aid of the first approximation theory of Gauss and Abbe; for as all the rays from any one object-point must pass very nearly through the conjugate image-points, the latter may be definitely located by the intersection of any two suitably selected rays starting from the object-point, and any one single ray traced from the object-point must be a geometrical locus of the image.

Referring to the accompanying diagram which represents an object OO_1 observed



through an objective of which only the outside surfaces are indicated, and through an eyepiece of the usual Huygenian type, it is assumed that the instrument has been focussed so that the rays from the axial object-point O, if not intercepted, would produce a primary image at F_e, the first or anterior principal focus of the eyepiece, and that the further refraction of these rays by the eyepiece lenses would therefore lead to their emergence from the eye-lens as a bundle of parallel rays which would be interpreted by the observer as coming from a very distant virtual image-point. To determine the image of the extra-axial object-point O_1, a ray

starting from this point in a direction parallel to the axis of the instrument is selected. By the general theory of lenses referred to, such a ray will emerge from the objective as if it had proceeded without deviation to the second or upper principal plane P'_0 of the objective and had there been bent so as to go through the second or upper focal point F'_0 of the objective, the axial distance from P'_0 to F'_0 representing the equivalent focal length f'_0 of the objective. This ray represents a geometrical locus of the image, and as a perfect objective would give an image at right angles to the optical axis of the similarly orientated object OO_1, $F_eO'_1$ must be the primary image, and similar triangles with a common corner at F'_0 at once give its magnification as

$$M_o = -\frac{F'_oF_e}{f'_o}, \quad . \quad . \quad . \quad (a)$$

the minus sign being added in accordance with convention because the image is inverted.

F'_oF_e, which is thus disclosed as a very important dimension, is known as the "optical tube-length" of the microscope, to distinguish it from the usually not very different "mechanical tube-length," which is measured from the lower to the upper extremity of the actual metal tube of the instrument.

Equation (a) supplies the means of determining the equivalent focal length f'_o, and as this is one of the chief characteristics of an objective, the method must be briefly described. The actual size of the primary image can obviously be read off if the image is received *directly* on a divided scale or "eyepiece micrometer." If such a micrometer, of Ramsden form, is available, it can be used as it is. Usually, however, microscopes have only Huygenian eyepieces and a simple glass-disc micrometer which can be laid upon the internal field-diaphragm of any of these. In such a case the field-lens of the eyepiece to be employed *must* be removed; this will reduce the field and will cause a great loss of sharpness in the extra axial images, but this must be put up with, as the field-lens, if left in position, would greatly reduce the size of the primary image and would therefore utterly falsify the results. By observing a "stage micrometer" divided in the same unit of length as the eyepiece micrometer, M_0 may then be measured at any extension of the draw-tube. A first measurement should be made with the draw-tube completely pushed in; let the result be

$$M_o^{(1)} = -\frac{F'_oF_e}{f'_o}.$$

The right side of this equation contains two unknown quantities and cannot therefore be solved for f'_0. But if now a second measurement of M_0 is made with extended tube-length, F'_oF_e will obviously be increased by the amount of the extension D which can be measured by an ordinary scale, and a second equation will be obtained,

$$M_o^{(2)} = -\frac{F'_oF_e + D}{f'_o},$$

and combination of this (by subtraction) with the equation for $M_o^{(1)}$ gives at once

$$f'_o = -\frac{D}{(M_o^{(2)} - M_o^{(1)})}.$$

When the equivalent focal length is thus determined, the location of F'_0 may be ascertained by measuring directly the location of the eyepiece micrometer with reference to the upper end of the tube, and by calculating F'_oF_e from the equation of the first observation.

The location of the upper focal plane with reference to the lower end of the tube varies considerably in different objectives, especially in those of low power. This explains why objectives of the same equivalent focal length may give decidedly different magnifications when exchanged on the same mechanical tube-length. Opticians frequently misstate the focal length of low-power objectives deliberately in order to make the magnification produced by them agree with what is traditionally expected from given focal lengths. There is also a traditional custom of assigning to ordinary objectives of high power a longer focal length than they really possess—a nominal $\frac{1}{12}$ inch being usually a $\frac{1}{14}$ inch; disagreements up to 20 per cent between the directly measured and the stated focal length need not therefore be taken as indicative of errors in the measurements.

The eyepiece magnifies the primary image just as an ordinary magnifying glass magnifies a real object, and M_e may therefore be taken as equal to the quotient of the adopted distance of distinct vision (conventionally 250 mm. or 10 inches) by the equivalent focal length of the eyepiece; f_e may be determined by the method already described by temporarily attaching the eyepiece to the lower end of the microscope tube, or any of the usual optical bench methods may be used. The total magnification of the instrument will then be $M = M_0 \times M_e$. The total magnification may be determined directly by projecting the image of a stage micrometer upon a scale at the conventional distance of distinct vision from the eyepiece, or more strictly from the bright disc of light ("Ramsden circle") seen a little above the eye-lens; this may be done visually by looking with one eye at the image of the stage micrometer and projecting the latter by binocular vision upon a scale held at the conventional distance from the other eye, or a real image of the stage micrometer may be projected upon a white surface and measured. A very clever simplification of this process is due to Mr. E. M.

Nelson; it consists in determining once for all the apparent diameter of the field-stop of each eyepiece when projected at the distance of distinct vision by one of the methods already referred to. The total magnification of any objective at any tube-length is then obtainable by a single observation by noting how many divisions of a stage micrometer are embraced by the field of the eyepiece in use. Thus, supposing that the actual diameter of the field of the eyepiece had been found as 162 mm. at the conventional distance of 250 mm., and that a certain optical combination was found to cover the diameter of this field with ·257 mm. of a stage micrometer, the total magnification would be

$$\mathrm{M} = -\frac{162}{\cdot 257} = -631 \text{ times.}$$

As owing to the very sensible distortion of all eyepieces the magnification is not a very accurately defined number, this process is amply sufficient and is certainly by far the most convenient.

Abbe preferred a totally different way of viewing the magnifying action of the compound microscope. Referring to the ray traced in the diagram from the extra-axial object-point O_1, it is obvious that if the eye were placed at F'_o the objective would act as an ordinary magnifying glass (provided the focus be slightly adjusted) and present the object under an angle U_p. Therefore Abbe called the ratio of the distance of distinct vision to the equivalent focal length of the objective the "initial magnifying power" of the objective and used this number as a constant of the objective. Following the ray from O_1 further, it is seen that it emerges from the eyepiece at an angle U'_p with the axis; therefore the object is seen through the whole microscope under this increased angle, and looking upon this increase of the angular subtense of the image as the characteristic function of the eyepiece, Abbe called the ratio $\tan U'_p/\tan U_p$ the "angular magnification of the eyepiece" and determined the total magnification as the product of the initial magnification of the objective and of the angular magnification of the eyepiece. If the numbers are correctly determined, the resulting magnification is necessarily the same as that found by the older and more familiar method. The data supplied with the Zeiss apochromatic objectives and their compensating eyepieces are based on this system, which has been copied by other makers of lens systems of these types.

On the older and preferable system the change of magnification due to change of tube-length is attributed to the objective; on the Abbe system it falls upon the eyepiece, the 12-times compensating eyepiece for the Continental tube-length being, for example, optically identical with the 18-times eyepiece for the English tube.

The Gauss theory of lenses may be applied to the microscope as a whole by tracing parallel bundles of rays in both directions through the entire instrument and so locating its principal planes and focal planes. The compound microscope is then found to be equivalent to a concave lens of very short focal length, and this aspect supplies the simplest solution of the problem of visual depth of focus. By the Gauss theory a lens must have an equivalent focal length $f' = -\mathrm{L/M}$ in order to produce a magnification of M times at a distance L from its second focal plane, and the object must be placed at a distance $l = f'/\mathrm{M}$ from its first or anterior focal plane, in order to yield a sharp image. Elimination of f' from the two equations gives $l = -\mathrm{L/M}^2$. If L is put as the conventional least distance of distinct vision to which the usual magnification numbers are referred, *i.e.* L = −250 mm., the distance of the object from the first focal plane must be $l = -250/\mathrm{M}^2$. The normal eye can be accommodated for any distance up to infinity, and can therefore see objects *in* the first focal plane of the complete microscope. Therefore l represents the total range of distance of the object accessible to a normal eye without change of focal adjustment of the instrument. The depth of focus (as it is usually called) due to accommodation of the eye is therefore $= 250/\mathrm{M}^2$, or

at magnification 10: depth of focus 2·5 mm.,
at magnification 100: depth of focus ·025 mm.,
at magnification 1000: depth of focus ·00025 mm.

At high magnifications the depth of focus is thus shown to be an evanescent quantity, and the extremely high demands by microscopists as to delicacy and reliability of the fine adjustment of the focus is fully justified. It should be added that if the object is embedded in a medium denser than air the depth of focus is increased in direct proportion to the refractive index of the medium. A second part of the focal range of optical instruments, due to the finite wave-length of light, will be referred to subsequently. It is not sufficiently large to modify seriously the requirements as to delicate focal adjustment. On the other hand, it represents the only available latitude of focal adjustment when the microscopic image is projected upon a screen or upon a photographic plate.

The most important property of a microscope is its resolving power. The naked normal eye can see small objects, such as points or parallel lines when separated by a distance from centre to centre of about ·1 mm. The microscope greatly diminishes this least distance, and it soon became a matter of great

interest to define this increased resolving power, to discover the laws on which it depends, and to ascertain whether there is any definite limit to it. The most obvious conclusion from geometrical optics that this ought to be a mere question of magnifying delicate detail up to the size which the normal eye can appreciate was soon realised as being quite wrong. It was found that any given optical combination yielded additional detail only up to a certain moderate magnifying power, beyond which the image became coarser or more fuzzy, without disclosing any additional detail. Carefully conducted experiments then led to the conclusion that the resolving power of well-corrected microscope objectives grew very nearly in direct proportion to the angle between the extreme marginal rays which could enter an objective from any one object-point, and this angle (twice the angle U in the diagram) therefore became the recognised criterion of resolving power, and was adhered to until about 1873. In that year Abbe published the first short account of his researches on microscopic vision, which were based on the diffraction of light produced by the delicate structure of microscopic objects. He proved that the resolving power is strictly proportional to the *sine* of half the angle which up to then had been accepted as the criterion, but is further proportional to the refractive index of the medium in which that angle is measured. He introduced the now universally accepted term "Numerical Aperture" for the true measure of resolving power thus discovered and gave it the symbol NA, which is therefore defined by

$$\mathrm{NA} = \mathrm{N} \cdot \sin \mathrm{U},$$

if U is the angle of the most oblique ray which can enter the microscope objective and N the index of the medium in which the angle is measured. Abbe also introduced a convenient direct-reading instrument, the "Apertometer," by which the numerical aperture can be determined. In the case of "dry" objectives, *i.e.* those in which a layer of air intervenes between the object and the front lens, the NA may be accurately determined without any instrument by placing the tube of the microscope in a horizontal position, setting up two candles at a distance of about 50 cm. from the objective and moving them at right angles with the optical axis until the images of the flames are on the point of disappearing at opposite edges of the clear aperture of the objective under test as seen by the eye looking down the tube of the microscope. The triangle formed by the two flames and the focal point of the objective then obviously has the angle 2U at the latter point, and U can be calculated by measuring the sides of the triangle.

The resolving power resulting from the Abbe theory for regularly spaced lines or rows of dots is expressed by the formula

$$d = \frac{\frac{1}{2}\text{ wave-length}}{\mathrm{NA}},$$

in which d is the centre-to-centre distance of the lines or rows of dots and the wave-length is the usual tabulated one, measured *in air*. As NA = N . sin U, and as sin U cannot exceed unity, the formula shows at once that the only possibilities for increasing resolving power beyond that usually realised lie in either the use of light of very short wave-length (ultra-violet), or in embedding the object in a medium of the highest possible refractive index and designing the objective so as to admit the widest possible cone of rays in the embedding medium. Both possibilities have been explored by Zeiss, the first by microscopes containing only quartz and fluorspar lenses and working with nearly monochromatic ultra-violet light, the second by a few immersion objectives using monobromide of naphthalene (N = 1·65) as immersion and embedding medium.

II. Optical Design of Microscope Objectives

§ (1) Historical Notes.—Although John Dollond had discovered in 1757 how chromatically and spherically corrected telescope object-glasses can be produced by combining suitably curved crown and flint lenses, opticians did not succeed in usefully applying this principle to the small lenses required for the microscope until some fifty years later. The early small achromatic lenses were invariably composed of a nearly equiconvex crown lens and a nearly plano-concave flint lens and were at first not cemented together. Although the microscopic object is always located at the shorter conjugate distance, the telescopic practice of turning the convex side of an achromatic lens towards the object was mechanically copied, with the result that even the achromatised lens had to be severely stopped down in order to reduce spherical aberration and coma to tolerable magnitude. Charles Chevalier of Paris first pointed out the absurdity of this procedure, and a very considerable advance immediately followed the simple reversal of the lens. He also was one of the first opticians who cemented the two components together and thus greatly reduced the difficulty of both making and mounting the lenses with sufficient precision. A further extension of the power of microscope objectives was secured by placing several achromatic lenses one behind the other, but without any definite principle as to relative power and separation of the components.

A most important advance in this respect

is due to a highly gifted English amateur, J. J. Lister (father of the late Lord Lister), who discovered in 1830 that every achromatic lens of the externally plano-convex form, which at that time was exclusively used for microscope objectives, has two pairs of conjugate points for which it is free from spherical aberration. Purely by wonderfully accurate experimental trials on the microscope he located one of these pairs of "aplanatic points" as realised when the lens produces a real inverted image of some definite magnification. Having adjusted the instrument for the corresponding tube-length he then tried another lens of the same type placed at various distances in front of the first (between it and the object), and found that a distance could be found at which complete freedom from spherical aberration of the combination was again secured. Obviously the front lens then produced a sharp and erect virtual image of the object at the first aplanatic point of the back lens. The doublets produced in this way not only had the advantage of freedom from spherical aberration at a much larger aperture than had previously been found attainable, but could also be freed from coma in the outer part of the field by selecting a suitable ratio between the focal lengths of the components, and the field was very much flatter than with the close combinations previously used. The type originated by Lister is in fact so excellent that it is almost exclusively used to this day for the lower powers of the microscope; it must, however, be added that the freedom from spherical aberration of the two separated components is no longer adhered to, for it has been found that still better results, especially as regards the higher aberrations, are obtainable by combining a spherically under-corrected back lens with an over-corrected front lens. The Lister principle admits of repeated application, for a third lens may obviously be placed in front of a doublet so as to increase the magnification still further, and such triplets were very soon constructed under Lister's advice.

These triplets, however, have not survived. When the Lister type of doublet is found insufficient, which happens when the numerical aperture approaches the modest value of ·3, the optician now adopts a type originated by Amici, consisting of two over-corrected cemented back lenses and a simple thick plano-convex front lens. Whilst actually simpler than the Lister triplets, this type offers far superior possibilities of good correction of all the important aberrations and allows of raising the numerical aperture to ·8 or even ·85, which corresponds to the admission of a cone of rays from any one point of the object of an angular extent of 106° to 116°. The vast majority of ordinary high-power microscope objectives correspond to this Amici type, with occasional modifications such as a triple cemented back lens instead of the more usual binary form.

The final steps in the perfecting of the microscope objective are almost entirely due to Abbe, who was the first to apply rigorous computing methods to the problem.

§ (2) Ray-tracing Methods for Microscope Objectives. (i.) *Fundamental Formula.*—The angles of incidence and of obliquity of the rays passing through a microscope objective are so large—reaching 30° and more even in the lower powers—that analytical methods of approximation are almost useless for purposes of ray-tracing. This must be done by rigorous trigonometrical formulae in strict accordance with the law of refraction.

The formulae almost universally used for this purpose appear to be due to Bessel (Discussion of the Königsberg Heliometer, in *Astronomische Untersuchungen*, Königsberg, 1841), but the slightly modified form given below is that found in Steinheil and Voit's handbook of applied optics.

Referring to *Fig.* 1, let the ray entering a refracting spherical surface AP at P be defined

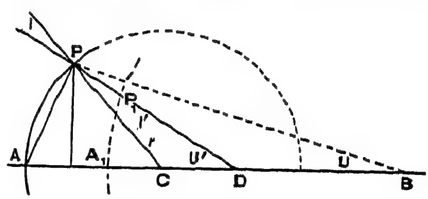


Fig. 1.

by its angle of obliquity U and by its intersection length AB, for which we introduce the symbol L. If C is the centre of curvature of the refracting surface and r its radius, simple trigonometry gives the angle of incidence CPB, with symbol I, by

$$\sin I = \frac{L-r}{r}\sin U. \qquad (1)$$

If we apply the same symbols, but distinguished by a dash, for the data of the refracted ray, the law of refraction next gives

$$\sin I' = \frac{N}{N'}\sin I. \qquad (2)$$

The angle at the centre of curvature ACP is exterior to both the triangle PCB formed by the incident ray and the triangle PCD formed by the refracted ray, hence

$$ACP = U + I = U' + I',$$

from which we obtain

$$U' = U + I - I'. \qquad (3)$$

The triangle PCD then gives the intersection length of the refracted ray

$$L' - r = \sin I' \frac{r}{\sin U'}, \quad . \quad . \quad (4)$$

$$L' = (L' - r) + r, \quad . \quad . \quad . \quad (5)$$

and the problem is solved.

An interesting and highly useful relation results by introducing the value of sin I given by (1) into (2), then this value of sin I′ into (4), namely,

$$(L - r)N \sin U = (L' - r)N' \sin U'. \quad . \quad (6)$$

The standard computing formulae may occasionally become objectionably inaccurate in the value of U′ found by them. U′ is found according to (3) by adding to the given U the change of direction of the ray $= I - I'$. If the angles of incidence and refraction are unusually large, as happens frequently at cemented contact surfaces and also in the front lenses of high-power objectives, then I and I′ are found by their sines or log-sines in a part of the table where a small change of the sine, and especially of its log, corresponds to a large change in the angle, and the latter becomes correspondingly uncertain. In such cases a more accurate value of $I - I'$ may be derived from the less accurate value of I and I′ obtained by (1) and (2) by a transformation of (2); by (2)

$$\frac{\sin I}{\sin I'} = \frac{N'}{N},$$

therefore $$\frac{\sin I - \sin I'}{\sin I + \sin I'} = \frac{N' - N}{N' + N},$$

and the formulae for sums and differences of sines give

$$\tan \tfrac{1}{2}(I - I') = \tan \tfrac{1}{2}(I + I')\frac{N' - N}{N' + N}. \quad . \quad (7)$$

The value of $I - I'$ obtained from this formula has, in the case of logarithmic calculation, only about one-third of the inaccuracy of the value obtained from (1) and (2) for glass-air refractions: in the case of contact surfaces the advantage is very much greater owing to $N' - N$ being then small compared to either N′ or N.

The accuracy of the L′ obtained from (5) will be insufficient when r is much larger than L′, because the quantity really determined directly from the data is $(L' - r)$, and if this is large, the usual logarithmic calculation will give only a few reliable decimal places and L′ will be uncertain accordingly. Moreover, there is no safeguard in the formulae against numerical errors, and as such an error at one surface will falsify the whole result, a reliable check is highly desirable. Both these desiderata are met in a most perfect manner by a simple formula proposed by Conrady: Draw the chord PA from the pole of the surface to the point of incidence. Triangle ACP is isosceles, and the angle at $C = U + I = U' + I'$, consequently the angle

$$APC = 90° - \tfrac{1}{2}(U + I) = 90° - \tfrac{1}{2}(U' + I').$$

The angles APB and APD between the chord and the incident and refracted ray exceed the angle APC by I and I′ respectively, and are therefore $90° - \tfrac{1}{2}(U - I)$ and $90° - \tfrac{1}{2}(U' - I')$ respectively. With these values the triangles APB and APD give at sight the check results

$$PA = L \sin U \sec \tfrac{1}{2}(U - I),$$

$$L' = PA \operatorname{cosec} U' \cos \tfrac{1}{2}(U' - I'). \quad (5) \text{ check.}$$

The value of L′ so found should agree within the limit of accuracy of the calculations with that by (5): as the value of L′ from the check formula is always more consistent with the corresponding value of U′, and usually also more accurate, the value of L′ found by the check should invariably be selected for retention. The value of PA found incidentally is also highly useful for other parts of the calculations of a lens system. It gives, for instance, convenient expressions for the rectangular co-ordinates of the point of incidence P with reference to the optical axis and the pole of the surface, for in the triangle APQ the angle at A is

$$= 90° - \tfrac{1}{2}(U + I) = 90° - \tfrac{1}{2}(U' + I').$$

Hence

$$X = PA \sin \tfrac{1}{2}(U + I), \quad Y = PA \cos \tfrac{1}{2}(U + I). \quad (8)$$

A still more convenient exact formula for X is obtained by continuing the trace of the refracting surface to its second intersection A′ with the optical axis and joining P and A′. The right-angled triangles PAQ and A′PA are similar and give

$$\frac{X}{PA} = \frac{PA}{2r},$$

or $$X = \frac{(PA)^2}{2r}. \quad . \quad . \quad . \quad (8^*)$$

This utility of PA for other purposes is the reason why the check formula is given in two parts; so computed it gives by its first part a value of PA which is proved to be correct by the verification of L′ and may therefore be used with absolute confidence.

(ii.) *Sign Conventions.*—The computing formulae (1) to (8) have been deduced for one particular case in which all the quantities entering into them were treated as positive. The formulae can be rendered valid for all possible cases by adopting appropriate sign conventions. Many of the sign conventions proposed in books on optics break down under certain special conditions. The following conventions are free from this objection, and also give the usual signs of focal lengths and other data, the optical axis being taken as horizontal:

(*a*) Intersection lengths L or L′ and radii of curvature are given the positive sign if the

intersection points or centres of curvature lie to the right of the refracting surface.

(*b*) The acute angle U between the optical axis and the incident ray is given the positive sign if a clockwise turn would carry a ruler through this acute angle from coincidence with the optical axis to coincidence with the ray.

(*c*) The signs of the starting values having been determined by the preceding rules, those of the derived quantities are fixed by the usual rules of algebra and trigonometry.

The angles entering into calculations of optical instruments are almost invariably acute, hence no ambiguity can ordinarily arise. The only exception is the angle U+I at the centre of curvature, which frequently slightly exceeds 90° in the front lenses of oil-immersion objectives. But as it is found by addition of two acute angles of definitely known sign, no doubt can arise in this case.

As the direction in which the light travels does not enter into the above sign conventions, the formulae can be used with equal confidence and without any modification for calculations in either direction, which at times proves very advantageous, especially in eyepieces.

In the case of microscope objectives the formulae are used exclusively for the tracing of rays from an original object-point on the optical axis of the system, and as the latter must be a perfectly centred one, the final as well as all the intermediate images will also lie on the optical axis. There is in fact complete symmetry with reference to the optical axis, and the result found for one ray at a given initial inclination U to the optical axis applies equally to rays of that inclination leaving the object point in any other azimuth with reference to the optical axis; in other words, the calculation for one ray covers a complete narrow circular zone of the system.

In this simplest case of the application of the general formulae the continuation of the tracing of a ray through a following surface also assumes a very simple form, for if A_1P_1 in *Fig.* 1 represents such a surface at axial distance d' from the preceding one, the obvious relations exist:

$$L_1 = L' - d', \quad U_1 = U', \quad N_1 = N', \quad . \quad (5^*)$$

and with these as starting values the ray can be traced through the new surface.

§ (3) Paraxial Rays.—Rays starting from the object-point at different values of U will in general give a different position of the respective final image-points, thus disclosing the longitudinal spherical aberration of the system, which must be sufficiently corrected if the system is to be a useful microscope objective.

The usual first approximation in the correction of spherical aberration consists in bringing together the image-points produced by the marginal zone and by the paraxial zone respectively. The computing formulae already given are easily adapted to the tracing of paraxial rays by noting that for these rays which pass very close to the axis all the angles will necessarily be very small; hence the difference between the sines and the angles themselves (measured in radians) becomes negligible, and cosines and secants become equal to one. If the paraxial angles and intersection lengths are distinguished by using small letters instead of capitals, the computing formulae therefore will be

$$i = u\frac{l-r}{r}, \quad . \quad . \quad . \quad . \quad (1p)$$

$$i' = i\frac{N}{N'}, \quad . \quad . \quad . \quad . \quad (2p)$$

$$u' = u + i - i', \quad . \quad . \quad (3p)$$

$$l' - r = i'\frac{r}{u'}, \quad . \quad . \quad . \quad (4p)$$

$$l' = (l' - r) + r, \quad . \quad . \quad (5p)$$

$$l_1 = l' - d', \quad u_1 = u', \quad . \quad . \quad (5p)^*$$

$$(l-r)N \, . \, u = (l'-r)N'u'. \quad . \quad (6p)$$

In the check formula the chord PA becomes indistinguishable from the ordinate of the point of incidence, with symbol y, and the check therefore is

$$y = l \, . \, u, \quad l' = \frac{y}{u'}. \quad . \quad (5p) \text{ check.}$$

Equation (7) becomes

$$i - i' = (i + i')\frac{(N' - N)}{(N' + N)}, \quad . \quad (7p)$$

and the equations for X give

$$x = \tfrac{1}{2}y(u + i) \, . \, \frac{y^2}{2r}. \quad . \quad . \quad (8p)$$

Strictly these formulae should be computed with a very small initial value of u. But inasmuch as (1p) to (7p) are all linear equations, it is evident that if a calculation first carried out with an appropriately small value of the initial u were repeated with $k \, . \, u$ as the initial inclination of the ray (k being any number whatever), all the other angles would come out at k times the first values, whilst the intersection lengths would be absolutely unchanged. As the latter are the quantities of chief interest, the initial u at the first surface of any one system may therefore be chosen quite freely, and in order to make the application of the sine condition as simple as possible it is most convenient to choose as the nominal value of the initial u the exact value of sin U for the corresponding marginal ray. Identity of the value found for the final u' of the system with the final value of sin U′ of the marginal ray then indicates fulfilment of the sine condition without a calculation of any kind.

It is, of course, necessary to bear in mind that these convenient large nominal values of the paraxial angles are fictitious when their absolute values are required for any special purpose.

§ (4) Special Formulae for Plane Surfaces.—Rays may be traced through plane surfaces at right angles to the optical axis by the general formulae by using a fictitious radius of curvature of such great length that the corresponding spherical surface only departs by a small fraction of a wave-length from a plane within the full aperture at which the surface works. But special formulae are far more convenient. Putting equation (1) into the form

$$\sin I = \sin U\left(\frac{l}{r} - 1\right),$$

it is seen that for $r = \infty$ we have $I = -U$. Equation (4) similarly shows that for a plane $I' = -U'$. With these special values of I and I′ equation (3) gives $\sin U' = N/N' \sin U$, and on introducing the special values into the check-formula a formula for L′ results, namely,

$$L' = L \tan U \cot U' = L\frac{N' \cos U'}{N \cos U},$$

by introducing the relation existing in this case between sin U and sin U′. The paraxial equations are easily deduced from the above and give the complete set of computing formulae:

$$I = -U, \quad I' = -U', \qquad \text{Pl}(1)$$

$$i = -u, \quad i' = -u', \qquad \text{Pl}(1p)$$

$$\sin U' = \frac{N}{N'} \sin U, \qquad \text{Pl}(2)$$

$$u' = \frac{N}{N'}u, \qquad \text{Pl}(2p)$$

$$L' = L \tan U \cot U', \qquad \text{Pl}(3)$$

$$l' = l \cdot \frac{N'}{N}, \qquad \text{Pl}(3p^*)$$

or

$$L' = L\frac{N'}{N} \cdot \frac{\cos U'}{\cos U}. \qquad \text{Pl}(3^*)$$

Pl(3*) is more easily computed than Pl(3), and gives a more accurate result if the angles are small. Pl(1) and (1p) do not enter into the calculation, but are important for certain general discussions.

§ (5) Spherical Aberration. (i.) *General Formulae.*—When a paraxial and a marginal ray are traced from an original object-point through a system, the difference $(l' - L')$ at any one surface indicates the longitudinal spherical aberration. Whilst the aberration is most easily found in this way it is obtained in a form which is highly misleading as to the real seriousness of the defect which is indicated, and even more so with reference to the real contribution to the total aberration by the separate surfaces. These undesirable peculiarities of the longitudinal spherical aberration become apparent and its laws are disclosed by finding a direct expression for it in terms of the computed angles and intersection lengths.

Standard equation (1) may be written

$$\frac{\sin U}{\sin I} = \frac{r}{L - r},$$

which gives by simple transformations

$$(a) \quad \frac{r}{L} = \frac{\sin U}{\sin I + \sin U},$$

and

$$(b) \quad \frac{L - r}{L} = \frac{\sin I}{\sin I + \sin U};$$

(*a*) may be further transformed into

$$\frac{r}{L} = \frac{\sin I + \sin U - \sin I}{\sin I + \sin U} = 1 - \frac{\sin I}{\sin I + \sin U},$$

and if this equation is multiplied throughout by N/r it gives

$$(c) \quad \frac{N}{L} = \frac{N}{r} - \frac{N}{r} \cdot \frac{\sin I}{\sin I + \sin U}.$$

Standard equation (4) is identical in form with (1), and treated in the same way as the latter gives

$$(d) \quad \frac{N'}{L'} = \frac{N'}{r} - \frac{N'}{r} \cdot \frac{\sin I'}{\sin I' + \sin U'}.$$

As (*d*) applies to the ray after refraction, the N′ sin I′ in its last term is equal—by the law of refraction—to N sin I in (*c*), hence we may combine (*c*) and (*d*) by subtracting the first from the second and treating N sin I as a common factor in the last terms, leading to

$$(e) \quad \frac{N'}{L'} - \frac{N}{L} = \frac{N' - N}{r} + \frac{N \sin I}{r} \cdot \left(\frac{1}{\sin I + \sin U} - \frac{1}{\sin I' + \sin U'}\right).$$

Taking $1/(\sin I + \sin U)$ outside the bracket, using (*b*), and in the bracket the formula for the sum of two sines, gives for the final term :

Final term

$$= \frac{N(L - r)}{rL} \cdot \left(1 - \frac{\sin \frac{1}{2}(U + I) \cos \frac{1}{2}(I - U)}{\sin \frac{1}{2}(U' + I') \cos \frac{1}{2}(I' - U')}\right).$$

By standard equation (3) $U + I = U' + I'$, hence the sines of half these angles cancel each other. Bringing the simplified bracketed term to a common denominator and applying the formula for the difference of two cosines next gives

$$\text{Final term} = 2\frac{N(L - r)}{rL} \times \frac{\sin \frac{1}{2}(I - U + I' - U') \sin \frac{1}{2}(I - U - I' + U')}{\cos \frac{1}{2}(I' - U')}$$

By elimination of U′ from the last numerator by (3): $U' = U + I - I'$, the final term takes its definitive form, and on introduction into (e) gives

$$\frac{N'}{L'} = \frac{N}{L} + \frac{N'-N}{r}$$

$$+ 2\frac{N(L-r)}{rL} \cdot \frac{\sin\frac{1}{2}(I'-U)\sin\frac{1}{2}(I-I')}{\cos\frac{1}{2}(I'-U')} \quad (9)$$

as a trigonometrically exact equation from which the laws of longitudinal spherical aberration can be deduced. If an object-point at any given distance, L, is considered, the first two terms on the right of (9) will be constant, hence variation of L′ for rays at different inclinations, U, can only result from the final term which is thus pointed out as the one which determines the spherical aberration. Its value depends chiefly on the two sines in the numerator which for moderate values of the various angles will, as was proved for the paraxial region, vary very nearly in direct proportion with the initial U of any one ray, and therefore also in proportion with the distance, Y, from the optical axis at which any ray penetrates the refracting surface. In first approximation the spherical aberration expressed by the last term therefore grows with the square of the aperture, and as a consequence becomes small of the second order for the paraxial region. For paraxial rays equation (9) therefore becomes

$$\frac{N'}{l'} = \frac{N}{l} + \frac{N'-N}{r}, \quad . \quad . \quad (9p)$$

which will be recognised as one of the best-known and most useful formulae for the tracing of paraxial rays.

By deducting (9*p*) from (9) and making very slight reductions a perfectly general and rigorous direct expression for the longitudinal spherical aberration at any spherical surface is obtained, namely

$$l' - L' = (l - L)\frac{N}{N'} \cdot \frac{l'L'}{lL}$$

$$+ 2\frac{N}{N'}(L-r)\frac{l'L'}{rL} \cdot \frac{\sin\frac{1}{2}(I'-U)\sin\frac{1}{2}(I-I')}{\cos\frac{1}{2}(I'-U')}, \quad (9^*)$$

in which the second term again represents the new spherical aberration produced by the refraction, whilst the first term shows how the spherical aberration $(l - L)$ of the incident rays combines with the new aberration; in other words, the first term expresses the addition theorem of longitudinal spherical aberration. It is seen that $(l - L)$ has to be multiplied first by N/N′, which may vary from about ·6 for a refraction from air into dense glass to about 1·7 for a refraction from dense glass into air. This product is further multiplied by the ratios of the conjugate distances of the paraxial and marginal ray respectively: each of these ratios may vary from zero if an incident ray parallel to the optical axis (l or $L = \infty$) is rendered convergent or divergent by the refraction, to infinite value if a convergent or divergent ray is rendered parallel to the optical axis by the refraction at the surface. The extraordinary extent to which the longitudinal spherical aberration is frequently found to fluctuate from surface to surface when common sense and experience render it obvious that the real departures of the rays from their ideal course is quite slight is thus fully explained. The second term in (9*) is easily seen to have similar misleading factors, so that the amount of the new aberration is also found in a form which requires careful scrutiny before making up one's mind as to whether the aberration is really serious or not. Strong evidence is frequently found in actual optical designs that even experienced practical computers are not sufficiently aware of the very indirect and misleading nature of the indications which the longitudinal spherical aberration supplies as to the real seriousness and magnitude of the defect at any one surface.

(ii.) *Conditions for Minimum Spherical Aberration.*—An extremely interesting and important question, especially in connection with microscope objectives, is whether the aberration produced by a spherical surface can become small or zero for pencils of finite aperture. One such case is quite obvious, for if the object-point is infinitely close to the refracting surface, then L will be infinitely small, rays at any practicable angle, U, will meet the surface very close to the optical axis so that the curvature of the surface will be inappreciable and its tangent plane may be substituted. In accordance with the formulae for plane surfaces, the L′ will then also be infinitely small and there is no scope for spherical aberration. This most obvious case is taken advantage of in all high-power microscope objectives to reduce the magnitude of the aberration which would arise at the first surface next the object if the latter were at a considerable distance, simply by making the "free working distance" very small.

Two other cases can be deduced from equation (9*). Discussion of its final term by which the new aberration is determined shows at once that this term will be zero if $L = r$, that is, for rays directed towards the centre of curvature. Under these conditions $L - r$ will be zero, which alone is sufficient. But in addition $\sin\frac{1}{2}(I - I')$ will also be zero, as for rays passing radially there is no deviation. It follows that for rays directed towards points close to the centre of curvature the aberration is small of the second order. This second case is also made use of in microscope

objectives of high power: it accounts for the very usual meniscus form of the lens next above the front lens. The second lens thus receives the diverging rays at small angles of incidence on its concave first surface and keeps the spherical aberration down to a manageable magnitude.

(iii.) *Aplanatic Refraction.*—In all cases not coming under the two which have been discussed all the factors entering into the aberration term of (9*) will have finite values, with the possible exception of $\sin \frac{1}{2}(I'-U)$. If this factor became zero the aberration would again vanish; that is the third and most important case of aplanatic refraction at a spherical surface.

$\sin \frac{1}{2}(I'-U)=0$ implies $I'=U$

and also $\sin I'=\sin U$.

By combining equations (1) and (2) it is easily found that quite generally

$$\sin I'=\frac{N}{N'}\,\frac{L-r}{r}\sin U.$$

If, as in the special case under discussion, $\sin I'=\sin U$, the general equation takes the form

$$1=\frac{N}{N'}\,\frac{L-r}{r},\quad\text{or}\quad\frac{L-r}{r}=\frac{N'}{N},$$

from which follows

$$\frac{L}{r}=\frac{N'+N}{N},\quad\text{or } L=r\frac{N'+N}{N}.$$

If L satisfies this equation there will be no spherical aberration, no matter how large the angles may be, right up to grazing incidence or emergence of the rays. As refractive indices are always positive, the case can only arise when L and r have the same sign, which implies that the object point lies on the concave side of the refracting surface. For refraction from air ($N=1$) into glass (say, $N=1{\cdot}5$) the equation gives $L=2{\cdot}5r$, so that the diagram used in deducing the standard computing formulae approximately represents this third case of aplanatic refraction. Several other interesting relations apply in this case. By (3) $I'-U=I-U'$, hence the case also implies $I=U'$, and if this is followed up by the aid of (2) and (4) in the same way as before it leads to

$$L'=r\frac{N'+N}{N'},$$

which for the numerical example already used gives $L'=1{\cdot}667r$. Combination of the equations for L and L′ gives the additional interesting relations

$$\frac{L'}{L}=\frac{N}{N'}\quad\text{and}\quad\frac{1}{L}+\frac{1}{L'}=\frac{1}{r}.$$

Finally, the two conditions to be fulfilled,

$$\sin I'=\sin U$$

and $$\sin I=\sin U'$$

give, on dividing the second by the first,

$$\frac{\sin U'}{\sin U}=\frac{\sin I}{\sin I'}=\frac{N'}{N}$$

by the law of refraction.

Therefore $\sin U'/\sin U=N'/N=$ a constant for the surface, and this supplies the highly important further information that the optical sine condition is fulfilled in this third case of aplanatic refraction. The two conditions to be fulfilled, $\sin I'=\sin U$ and $\sin I=\sin U'$, also show, inasmuch as the angle of incidence in the less dense medium may reach but cannot exceed 90°, that the angle of convergence U or U′ in the denser medium has the same range, that is, this case permits of dealing with cones of rays up to an included angle of 180° between opposite extreme marginal rays in the denser medium quite independently of the values of N and N′. It is this unconditional freedom from spherical aberration for cones of rays of the widest possible angular extent which renders this third case of aplanatic refraction so extremely valuable in the design of microscope objectives. It may indeed be stated that microscope objectives of high aperture are possible only by making use of this property of spherical surfaces. The curved side of the front lenses always corresponds fairly closely to the conditions which have been deduced. If the conjugate points are a little nearer to the refracting surface than is required for exact aplanatism, then a very considerable collective refraction can be secured accompanied by spherical over-correction such as is usually only obtainable at the expense of a dispersion of the rays. Front lenses therefore depart from the exact theoretical position of the aplanatic points more frequently in this sense than in the opposite sense of too great a distance of the conjugate points from the refracting surface.

It is an obvious conclusion drawn from the above discussion that if the aberration accompanying a given deflection $(I-I')$ is desired to be large, as most often happens at the dispersive contact surfaces at which over-correction is produced, then a large value of $(I'-U)$ must be aimed at by having the conjugate points as far as possible away from the aplanatic position. Hence for conjugates lying at the concave side of a surface their distance should be a large multiple of the radius; but still higher aberrations will be secured by having the conjugate points lying on the convex side of the surface, which is in fact the predominating practice.

§ (6) Zonal Spherical Aberration. *Expression as a Series.*—It has been proved that

the longitudinal spherical aberration grows, in first approximation, with the square of the aperture of any given lens system. The form of equations (9) and (9*), however, makes it obvious that there must be higher terms in their development in series. The type of this series may be fixed without direct mathematical investigation by the simple consideration that on account of the symmetry of a centred system with reference to its optical axis a ray from an axial object-point entering at the same angle below the optical axis as another above the axis must suffer the same longitudinal aberration as to both magnitude and sign. Two such corresponding rays will have the ordinate Y of the point of incidence as well as the chord PA of the same numerical value but of opposite sign, whence it follows at once that the series giving the longitudinal aberration in terms of powers of Y or of PA can only contain even powers of these as well as of any other direct measures of the aperture; therefore the longitudinal aberration LA′ must be represented by an equation of the form

$$LA' = c_1Y^2 + c_2Y^4 + c_3Y^6 + \text{etc.}$$

In systems of slight curvature such as astronomical object-glasses of the usual types, only the first term is sensible. Such systems can have their spherical aberration corrected in a practically perfect manner *on paper*. The very sensible residues of aberration frequently found by direct tests (such as Hartmann's) in such systems are due to a slight extent to departures from the prescribed radii and thicknesses, to a greater extent to departures from true spherical form of the polished surfaces due to elastic yielding of the lens and tools in the process of grinding and polishing, but chiefly, especially in large sizes, to want of homogeneity in the glass which at times even affects microscope objectives. Empirical "figuring" of a suitable surface is the only remedy for these technical defects, which cannot be either estimated or allowed for in the calculation of a lens system.

In lens systems of somewhat bolder design the first two terms of the LA′ series represent the longitudinal aberration with sufficient accuracy. If such a system is spherically corrected for a particular zone of its aperture (usually the extreme marginal zone) the terms in Y^2 and Y^4 must be numerically equal but of opposite sign for that zone, whilst there will be residuals of aberration of the sign of the Y^2 term between the paraxial and the corrected zone, and of the sign of the Y^4 term beyond the corrected zone, the former attaining a maximum value at $\sqrt{\frac{1}{2}}$ of the diameter of the corrected zone. Hence the rule that if the presence of zonal aberration is to be ascertained a third ray (in addition to the paraxial and marginal ones) should be traced through the system at ·7071 of the aperture selected for the marginal ray.

In systems of deep curvature, practically in all microscope objectives exceeding an NA of about ·3, still higher terms become sensible in rapid succession as the aperture increases, and a higher state of spherical correction must then be aimed at. In purely geometrical ray-tracing this is attained by demanding that the rays passing through three selected zones of the aperture shall be brought to a common focus, and it is usual to choose the paraxial, marginal, and the ·7071 zones.

If only the first three terms of the aberration series were of sensible magnitude this would lead to perfect spherical correction for the whole aperture. In reality the higher terms follow so closely on each other's heels that the tracing of intermediate rays again reveals secondary zonal variations. But there is no systematic method for the removal of these higher zonal variations.

§ (7) The Physical Aspect of Spherical Aberration. (i.) *General Consideration.*—It will have been gathered from the preceding geometrical discussion of the problem of spherical aberration that its chief weakness lies in the method of measuring the magnitude of the defects which the longitudinal spherical aberration indicates. It is usually attempted to estimate the true magnitude by discussion of the series for LA′ with a view to determining the smallest diameter of the confused pencil of rays resulting from the aberration. It is then assumed that the focus of the instrument will be adjusted so as to throw this smallest possible geometrical image upon the retina of the eye (or upon the photographic plate) and that each object-point will be seen or photographed as a dot of the size of this "circle of least confusion." When the spherical aberration is of very considerable magnitude, these assumptions are very nearly correct. Cases of that kind, however, are of extremely rare occurrence in practically useful optical instruments. Telescopes and microscopes of the present day are expected to realise the full theoretical resolving power of a perfectly corrected instrument as defined by the integration of the light effect at a perfect focus in accordance with the undulatory theory of light. As the limit of defining power so determined is decidedly stringent, only small amounts of aberration can be tolerated. In this, practically the only interesting case, the geometrically determined circle of least confusion becomes absolutely useless as a measure of the size of the actual image of a point: as will be shown, the geometrical circle of least confusion may be four or more times the size of the image really seen by the eye before any sensible deterioration due to spherical aberra-

tion sets in. It might be argued that this must surely be in favour of adhering to the geometrical measure, inasmuch as an instrument satisfying the geometrical limit would prove quite perfect when submitted to actual test by light of finite wave-length. The answer to such an argument is that in order to reduce his circles of least confusion to the size of the physically determined image of a point, the purely geometrical designer would have to stop at a very much reduced aperture of a system of any given number of component lenses. As the true resolving power is proportional to the aperture, he would thus invariably produce a system of lower resolving power and with a smaller light gathering power than could be produced with the same number of components by taking the finite wave-length of light into account in determining the permissible aperture. Adherence to the geometrical method must therefore prove a severe, if not a fatal, handicap whenever the best possible result is to be secured with a given number of components.

Physically considered, a lens system changes the curvature of the light-waves passing through it: if perfect the system turns the spherical waves sent out by any object-point into truly spherical waves converging towards the conjugate image-point; consequently the light sent out at a given instant from the object-point towards every part of the clear aperture would a little later arrive absolutely simultaneously at the conjugate image-point and would produce a maximum brightness of the image owing to the total absence of differences of phase. The simultaneous arrival at the image-point of light traversing all parts of the clear aperture is therefore the physical definition of a perfect optical system. Inasmuch as the rays of geometrical optics are normals of the corresponding wave-surfaces, this physical definition of a focus is in the case of perfect systems identical with the geometrical one; for if the waves converging towards the image-point are spherical, then their normals, the "rays," will be radii and will therefore meet at the image-point.

(ii.) *Physical Theory.*—But whilst the geometrical theory claims a perfect point-focus, the undulatory theory merely demonstrates that there is a maximum of intensity at that point, because the light reaches it without any differences of phase. In accordance with the principle of Huygens as extended by Fresnel, even a truly spherical wave produces a light-effect outside the cone of "rays" to which it corresponds. If, in *Fig.* 2, O represents an object-point sending out spherical waves like W and if the lens system changes these waves into a form like W′ converging more or less perfectly towards an image-point O′, then, according to the principle referred to, the light-effect at any point whatever beyond W′ can be determined by considering each surface element of W′ as a luminous point sending

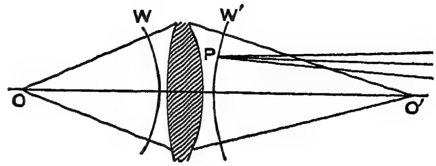


FIG. 2.

out spherical waves and by integrating the combined effect at any point of these elementary waves over the whole surface of W′. G. B. Airy [1] carried out this integration for the intensity distribution produced by a perfect lens system in the plane of the geometrical image and arrived at the well-known "spurious disc" surrounded by concentric diffraction rings of rapidly diminishing intensity. In this simplest case the light from all points of the emerging wave W′ will arrive at O′ in the same phase. If there is spherical aberration present, then the emerging waves will be distorted and will no longer be of the ideal spherical form, and there will not be a definite point like O′ at which all light arrives in the same phase. In order to discuss these cases (strictly speaking, the only ones which are of practical interest as perfect spherical correction is utterly impossible) it is necessary to determine the phase relation with which light arrives at any assumed position of the image-point. This is done by finding an algebraical expression for the total length of optical path from an object-point to the assumed image-point, in the first instance for a single spherical refracting surface.

In *Fig.* 3, let AP represent the trace of a spherical refracting surface with centre at C,

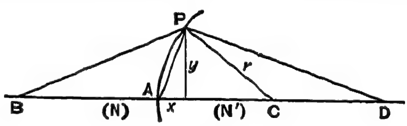


FIG. 3.

separating a medium of index N to its left from a medium of index N′ to its right; let B be the luminous object-point at distance l from the pole A, and D at distance l' from the pole the point for which the differences of optical path are to be determined, using the axial path B A D as the reference standard. Physically the refractive index represents the ratio of the velocity of light in empty space to that in the given medium, hence a path l in a medium of index N takes as long to

[1] *Camb. Phil. Trans.*, 1834.

traverse, or is optically equivalent to, a path Nl in empty space. Reducing all paths to their equivalents by this relation, there results —bearing in mind that l in the diagram is negative in accordance with the adopted sign convention—

Equivalent axial path $= N' . l' - N . l$,
Equivalent marginal path $= N' . PD - N . BP$.

Introducing the rectangular coordinates x and y of point P the second equation becomes

$$\text{E.m.p.} = N'\sqrt{(l'-x)^2+y^2} - N\sqrt{(l-x)^2+y^2},$$

in which the roots must obviously be taken with the signs of l' and l respectively. For a spherical surface $y^2 = 2rx - x^2$, hence

$$\text{E.m.p.} = N'\sqrt{l'^2 - 2x(l'-r)} - N\sqrt{l^2 - 2x(l-r)}.$$

The difference axial—marginal path will determine by how much the axial path is longer than the marginal path; or more conveniently put, it gives the lead of the marginal light. Introducing the symbol OPD for this optical path difference, it is found as

$$\text{OPD} = N'l'\left(1 - \sqrt{1 - 2x\frac{l'-r}{l'^2}}\right) - Nl\left(1 - \sqrt{1 - 2x\frac{l-r}{l^2}}\right).$$

Putting $p = x(l-r)/l^2$ and $p' = x(l'-r)/l'^2$ and developing the roots by the binomial theorem, the result is

$$\text{OPD} = N'l'[1 - (1 - p' - \tfrac{1}{2}p'^2 - \tfrac{1}{2}p'^3, \text{ etc.})] - Nl[1 - (1 - p - \tfrac{1}{2}p^2 - \tfrac{1}{2}p^3, \text{ etc.})],$$

or, differently ordered,

$$\text{OPD} = N'.l'p' - N.lp + \tfrac{1}{2}(N'.l'p'^2 - N.lp^2) + \tfrac{1}{2}(N'.l'p'^3 - N.lp^3), \text{ etc.},$$

and this is valid for any position on the optical axis of the assumed image-point D.

If the values of p and p' are reintroduced and x, in accordance with the last of equations (8), is put equal to $(\text{PA})^2/2r$, the value of OPD becomes

$$\text{OPD} = \frac{(\text{PA})^2}{2r}\left(N'\frac{l'-r}{l'} - N\frac{l-r}{l}\right) + \tfrac{1}{8}(\text{PA})^4\left(\frac{N'(l'-r)^2}{r^2l'^3} - \frac{N(l-r)^2}{r^2l^3}\right) + \tfrac{1}{16}(\text{PA})^6\left(\frac{N'(l'-r)^3}{r^3l'^5} - \frac{N(l-r)^3}{r^3l^5}\right) + \text{etc.} \quad (10)$$

Evidently the first term in $(\text{PA})^2$ will disappear if the bracketed factor becomes zero. But

$$N'\frac{l'-r}{l'} - N\frac{l-r}{l} = 0$$

can be easily converted into

$$\frac{N'}{l'} = \frac{N'-N}{r} + \frac{N}{l},$$

an equation already deduced (see (9p)) as the condition fulfilled by the location of the paraxial geometrical image of the object-point. The difference of optical paths with which a ray at finite distance from the optical axis reaches the paraxial focus is therefore

$$\text{OPD}_p = \tfrac{1}{8}(\text{PA})^4\left(\frac{N'(l'-r)^2}{r^2l'^3} - \frac{N(l-r)^2}{r^2l^3}\right) + \tfrac{1}{16}(\text{PA})^6\left(\frac{N'(l'-r)^3}{r^3l'^5} - \frac{N(l-r)^3}{r^3l^5}\right), \text{ etc.} \quad (10^*)$$

This equation obviously gives the spherical aberration in terms of differences of optical paths. The latter will be obtained in wave-lengths by multiplying the right side by the number of wave-lengths contained in one unit of length (approximately by 2000 if the millimetre is used or by 50,000 if the inch is adopted) and is seen to grow with the fourth and higher even powers of the aperture, whereas the longitudinal aberration as defined geometrically grows with the square of the aperture.

The differences of phase therefore increase very much more rapidly than the equivalent longitudinal aberration when they once have reached sensible magnitude; on the other hand, the differences of phase will remain insensible in a zone surrounding the optical axis which is much larger than that in which the longitudinal aberration remains small.

As the OPD found for a ray at a surface measures the lead which it has gained with reference to the corresponding axial ray, it is evident that the addition theorem of the OPD for a number of surfaces is the simplest possible one, namely, simple algebraical addition of the separate values found for each surface. This is, however, strictly correct only if equation (10) is used and if the l' introduced into it is the true intersection length of the marginal ray as found trigonometrically. The reason for this restriction is that in the propagation of waves of an extent which is a large multiple of a wave-length the energy represented by the vibrations travels in the direction of the geometrical rays and must therefore be traced along the latter. The direction towards the paraxial focus differs from that towards the actual intersection point of a ray at finite angles by the amount of the angular value of the spherical aberration, and a correction is called for which, although small and easily determined, is a troublesome addition to the numerical work and makes another method of calculating the OPD to be given presently decidedly more convenient.

The correction referred to brings in small additional terms of the 6th and higher even orders; it therefore does not alter the law of increase of the OPD with aperture, so that for any centred optical system the OPD as given by (10) will be expressed by an equation of the form

$$\text{OPD} = c_1Y^2 + c_2Y^4 + c_3Y^6 + \text{etc.},$$

for any assumed position of the image-point, and by an equation (10*) of the form

$$OPD_p = d_1 Y^4 + d_2 Y^6 + \text{ etc.},$$

for the paraxial image-point. The presence of a term in Y^2 for points other than the paraxial focus is of great importance in the discussion of the location of the best image and in laying down the limits of tolerance for residual spherical aberration. When the equations are written with undetermined coefficients c_1 etc., they hold for any reasonable measure Y of the aperture, as for instance the ordinate of the point of penetration of the ray through any one surface of the system, or the angle of convergence of the issuing or entering ray, for all these different measures are convertible into each other by series progressing in alternate powers of any given measure of the aperture, by reason of the symmetry of a centred system around the optical axis.

§ (8) Trigonometrical Calculation of the OPD.—Let a spherical wave OA, *Fig.* 4,

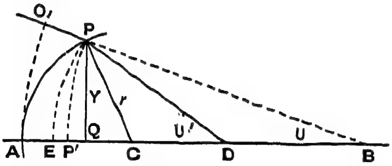


Fig. 4.

in the medium of index N to the left of the refracting surface AP, be converging towards B. While the marginal part of the wave is traversing the distance OP in the medium of index N, the axial part has entered the new medium of index N′ and the optical paths of the two parts are necessarily equal. As the marginal ray is refracted towards D in the direction given by the ordinary law of refraction a new spherical wave EP, with D as centre, would result if there were no spherical aberration. Therefore in this case the optical path AE would be equal to the optical path OP. Conversely, if the optical paths OP and AE are separately computed and found to differ, the difference will represent the OPD due to spherical aberration.

Describe an arc of a circle PP′ with B as centre cutting the axis in P′, and draw PQ perpendicular to the axis.

Then $\qquad OP = AP' = AQ - P'Q.$

Join AP and P′P. The triangles CAP and BPP′ are both isosceles. As the angle at C is $= U + I$ and the angle at $B = U$, it follows that angle $PAQ = 90° - \frac{1}{2}(U+I)$ and angle $PP'Q = 90° - \frac{1}{2}U$, whence putting $PQ = Y$

$$OP = Y(\tan \tfrac{1}{2}(U+I) - \tan \tfrac{1}{2}U).$$

The distance AE may be evaluated in the same way remembering that $PCA = U' + I'$ and $PDA = U'$, and we find

$$AE = Y(\tan \tfrac{1}{2}(U'+I') - \tan \tfrac{1}{2}U').$$

And the OPD at the marginal focus D is found by multiplying AE and OP by the respective indices and taking their difference. Thus,

$$OPD_m = N'.\, Y(\tan \tfrac{1}{2}(U'+I') - \tan \tfrac{1}{2}U')$$
$$- N\,.\, Y(\tan \tfrac{1}{2}(U+I) - \tan \tfrac{1}{2}U).$$

Whence on reduction we have

$$OPD_m = N'.\, Y \frac{\sin \frac{1}{2}I'}{\cos \frac{1}{2}(U'+I') \cos \frac{1}{2}U'}$$
$$- N\,.\, Y \frac{\sin \frac{1}{2}I}{\cos \frac{1}{2}(U+I) \cos \frac{1}{2}U}.$$

Multiplying numerator and denominator of the first term by $2 \cos \frac{1}{2}I'$, and of the second by $2 \cos \frac{1}{2}I$, the numerators of the fractions become sin I′ and sin I respectively, and as $N' \sin I' = N \sin I$ and $(U+I) = (U'+I')$ the equation becomes

$$OPD_m = \frac{N'.\, Y \sin I'}{2 \cos \frac{1}{2}(U+I)}$$
$$\left[\frac{1}{\cos \frac{1}{2}I' \cos \frac{1}{2}U'} - \frac{1}{\cos \frac{1}{2}I \cos \frac{1}{2}U}\right],$$

or on bringing the square bracket to a common denominator

$$OPD_m = \frac{N'.\, Y \sin I'}{2 \cos \frac{1}{2}(U+I)} \cdot \frac{\cos \frac{1}{2}U \cos \frac{1}{2}I - \cos \frac{1}{2}I' \cos \frac{1}{2}U'}{\cos \frac{1}{2}I \cos \frac{1}{2}U \cos \frac{1}{2}I' \cos \frac{1}{2}U'}.$$

The numerator of the second fraction can be further simplified by putting $\cos \frac{1}{2}U = \cos(\frac{1}{2}(U-U') + \frac{1}{2}U')$ and resolving, also putting $\cos \frac{1}{2}I' = \cos(\frac{1}{2}(U-U') + \frac{1}{2}I)$, which is obtained by transposing $U + I = U' + I'$, and resolving this. The numerator is then obtained in the form of four terms, two of which cancel each other whilst the other two combine and lead to the equation

$$OPD_m = \frac{N'.\, Y \sin I' \sin \frac{1}{2}(U-U') \sin \frac{1}{2}(I-U')}{2 \cos \frac{1}{2}U \cos \frac{1}{2}I \cos \frac{1}{2}(U+I) \cos \frac{1}{2}I' \cos \frac{1}{2}U'}.$$

It is more convenient for computation to replace the ordinate Y by the chord PA. Using $Y = PA \cos \frac{1}{2}(U+I)$ this gives

$$OPD_m = \frac{N'.\, PA\,.\, \sin I' \sin \frac{1}{2}(U-U') \sin \frac{1}{2}(I-U')}{2 \cos \frac{1}{2}U \cos \frac{1}{2}I \cos \frac{1}{2}I' \cos \frac{1}{2}U'}. \qquad (10^{**})$$

For a plane surface PA becomes identical with Y and as by Pl(1) $I = -U$ and $I' = -U'$, the equation for plane surfaces is

$$\text{plane}: \; OPD_m = \frac{N'.\, Y \sin U' \sin \frac{1}{2}(U+U') \sin \frac{1}{2}(U-U')}{2 \cos^2 \frac{1}{2}U \cos^2 \frac{1}{2}U'}, \qquad (10^{**})$$

in which $Y = L \tan U = L' \tan U'$.

These equations are more rapidly computed than the number of terms would suggest because they give the OPD directly, not as a small difference of two large numbers.

As it is hardly ever of interest to know the aberration at any one surface within less than 0·1 per cent, four-figure logs are quite sufficient and the cosines in the denominator rarely call for any interpolation, as the angles are usually small. The aberration at the final marginal focus of a system is strictly equal to the algebraical sum of the OPD_m values found for the separate surfaces because the ray is traced in this case along the path assigned to it by the law of refraction. There is therefore no correction of any kind to be applied to the result obtained for any one ray. On the other hand, should several rays, at different distances from the optical axis, be traced through a system, the OPD_m obtained for them will not be directly comparable, as each one will be referred to its own geometrical intersection point. General theoretical discussions are therefore preferably based on Equation (10).

The differences of optical paths are a direct measure of the distortion from true spherical form of the waves emerging from a lens system. For that reason they make possible a really valid discussion of the limits within which spherical aberration must be corrected if the full theoretical resolving and defining power of a lens system is to be realised. This discussion is of particular importance in the case of microscope-objectives on account of their deep curvatures and consequent large amount of higher aberrations : but the results are equally valid for any other kind of centred optical systems. The foundations for this treatment of the aberrations were laid by the late Lord Rayleigh.

§ (9) The Rayleigh Limit.—In 1879 Lord Rayleigh[1] arrived at the conclusion that an optical system would give an image only slightly inferior to that produced by an absolutely perfect system *if all the light arrived at the focus with differences of phase not exceeding one quarter of a wave-length*, "for then the resultant cannot differ much from the maximum." This paper ought immediately to have marked the beginning of a new epoch in the design and discussion of optical instruments. But the barren, purely geometrical treatment of the aberrations went on for many years, possibly because the conclusion was jumped at that equalisation of the optical paths from object-point to image-point (themselves rarely measuring less than eight inches) within one two-hundred-thousandth part of an inch was an impossible ideal. As will be shown, this is so far from being correct that in reality it is far easier to correct the aberrations within the Rayleigh limit than to bring the geometrical "rays" within a "circle of least confusion" of the size of the image which has for a long time been known as attainable. The adoption of the Rayleigh limit thus makes it possible considerably to increase the aperture of a lens system of given type and to come close to the full theoretical resolving power with systems which, judged geometrically, would appear hopelessly over- or under-corrected.

These remarkable and highly valuable facts are easily proved by determining the aperture at which the Rayleigh limit is reached in the presence of aberration and by establishing the relation between the physical and the geometrical measurement of the aberration.

If D_p in *Fig.* 5 is the paraxial focus of a lens system, then in the absence of aberration

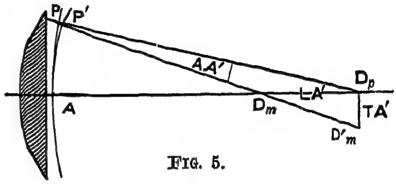


Fig. 5.

a train of spherical waves like PA would be converging towards this focus. Aberration will cause the actual waves to be distorted, and the OPD equations will determine directly the distance PP′ by which the distorted wave is separated from the corresponding ideal spherical wave. If the law of increase of the OPD with increasing aperture is known, the complete form of the distorted wave with reference to the ideal wave will be defined. In all optical instruments intended for high resolving power the angle of the cones of rays converging towards the final image point is always very small (much smaller than the exaggerated diagram suggests), the emerging waves rarely reaching an angular extent of even 5°. Under these conditions the difference between the arc AP and the ordinate Y of point P will be negligible and the differential coefficient $d(PP')/dY$ will be the measure in radians of the small angle between the ideal and the distorted wave and therefore also between their corresponding normals. The normal of the actual emerging wave represents by definition the true geometrical ray passing through any point P′. If D_m is the intersecting point of this ray, the angle $D_pP'D_m$, or the angular aberration of the ray, denoted by the symbol [AA′], is therefore defined by

$$[AA'] = \frac{d(OPD_p)}{dY},$$

no matter of what order the wave distortion may be, provided only that the angle of the cone of rays is reasonably small. If the ray at finite angle is produced to intersection at D'_m with a normal plane described through D_p, then the distance $D_pD'_m = TA'$ represents the transverse aberration and triangle $D_pP'D'_m$ gives with ample approximation

$$TA' = l'[AA'] = l'\frac{d(OPD)}{dY}.$$

[1] *Collected Papers*, i. 415-453; *Phil. Mag.*, Oct. 1879, viii. 261.

Finally the triangles $D_mD_pD'_m$ and D_mAP which are approximately similar give the longitudinal aberration $D_mD_p=LA'$ as

$$LA'=TA'\frac{L'}{Y}=\frac{l'L'}{Y}\,\frac{d(OPD_p)}{dY}.$$

These important relations between the differences of optical paths and the various measures of the corresponding geometrical aberrations may be collected in the form

$$\frac{d(OPD_p)}{dY}=[AA']=\frac{TA'}{l'}=LA'\frac{Y}{l'L'}. \qquad (11)$$

As in the case of the geometrical aberrations, so in that of the differences of optical paths the best image is usually found at some little distance from the paraxial focus. The effect of such a change of focus must therefore be determined.

Let D in *Fig.* 6 at a distance df from the paraxial focus D_p be a point for which the

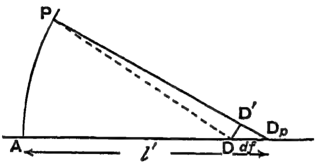


FIG. 6.

phase relation of light from P and from A respectively is to be determined. At D_p there will be a lead of the inclined ray by OPD_p. The axial path to D is shorter than that to D_p by df. The inclined path from P to D is shorter than that to D_p by the difference PD_p-PD.

In all cases of practical interest df is very small compared to DP, and the angle DPD_p measures a moderate number of seconds of arc. Under these conditions the difference between PD_p and PD will be sensibly equal to DD_p, the projection of DD_p upon PD_p. Consequently the light from P will arrive at D with a lag $=df(1-\cos DD_pD')=2df\sin^2\frac{1}{2}DD_pD'$.

Under the restriction to small angles of the cone of rays $\sin DD_pD'/2$ is sensibly equal to $Y/2l'$. Hence the phase relation at D will be

$$OPD=OPD_p-\tfrac{1}{2}df\left(\frac{Y}{l'}\right)^2, \quad . \quad (11^*)$$

in which df is treated as positive when lying towards the lens system, which is contrary to the usual sign convention but is in accordance with the general custom of treating spherical under-correction as positive.

§ (10) ABERRATION ON THE WAVE THEORY.—The principal cases of interest can now be discussed. (i.) *Ordinary or primary Spherical Aberration.*—The wave distortion for the paraxial focus will be represented by the first term of 10*, or

$$OPD_p=k\,.\,Y^4 \ (k \text{ a constant}),$$

or $$\frac{d(OPD_p)}{dY}=4kY^3=4\frac{OPD_p}{Y}.$$

By equations (11) these give the corresponding geometrical aberrations as

$$OPD_p=[AA']\,.\,\frac{Y}{4}=TA'\frac{Y}{4l'}=LA'\frac{Y^2}{4l'^2},$$

and if the last of these equivalents of OPD_p is put into (11*) the phase relation at any shifted focus is obtained as

$$OPD=\frac{Y^2}{4l'^2}(LA'-2df).$$

If $df=LA'$, that is, for the marginal focus, this gives

$$OPD_m=\frac{Y^2}{4l'^2}(-LA')=-OPD_p.$$

In the case of simple spherical aberration the axial and marginal light therefore meet with precisely the same difference of optical paths at the marginal as they do at the paraxial focus, only the sign of the difference is reversed.

If $df=\frac{1}{2}LA'$ or for the point exactly midway between paraxial and marginal focus, the difference of optical paths between axial and marginal light disappears. These relations become geometrically obvious if a diagram is drawn (*Fig.* 7). The true distorted wave represented by the thick curve bends away from the ideal wave for the paraxial focus in proportion to the fourth

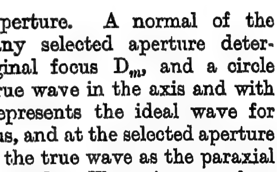


FIG. 7.

power of the aperture. A normal of the true wave at any selected aperture determines the marginal focus D_m, and a circle tangent to the true wave in the axis and with D_m as centre represents the ideal wave for the marginal focus, and at the selected aperture lies as far within the true wave as the paraxial ideal wave lies outside. The existence of an intermediate point D such that its ideal wave cuts the true one at the selected aperture is at once obvious. But it is also readily seen that this is the closest fitting spherical wave within the selected aperture, for if D were moved towards D_m the lump formed by the true wave between axis and margin would manifestly become higher, whilst if D were moved from the midway point towards D_p the intermediate lump of the true wave would diminish

to a less extent than the rapid growth of the projection of the marginal part of the true wave within the spherical wave under consideration. It may therefore be taken without algebraical proof (which, however, is easily supplied) that the residual differences of phase are smallest at the point midway between the paraxial and extreme marginal geometrical foci.

Their amount is easily determined. In the general equation

$$\mathrm{OPD}=\tfrac{1}{4}\left(\frac{Y}{l'}\right)^2(\mathrm{LA}'-2df)$$

LA′ represents the longitudinal spherical aberration for the zone of semi-aperture Y and by the law of simple spherical aberration varies with Y^2. Now let the full aperture be Y_1 and the corresponding spherical aberration LA'_1; then, by the law just referred to,

$$\mathrm{LA}'=\mathrm{LA}'_1\frac{Y^2}{Y^2_1},$$

and the equation for OPD becomes

$$\mathrm{OPD}=\frac{\mathrm{LA}'_1}{Y^2_1}\,\frac{Y^4}{4l'^2}-df\frac{Y^2}{2l'^2}.$$

This is easily proved to give a maximum value for $Y^2=dfY^2_1/\mathrm{LA}'_1$, and on putting this value of Y^2 into the equation for OPD it becomes

$$\mathrm{OPD}_{\mathrm{maximum}}=-\tfrac{1}{4}\frac{Y^2_1}{l'^2}\,\frac{df^2}{\mathrm{LA}'_1}.$$

For the best focus $df=\tfrac{1}{2}\mathrm{LA}'_1$, hence at mid-focus

$$\mathrm{OPD}_{\mathrm{maximum}}=-\tfrac{1}{16}\,\frac{\mathrm{LA}'_1Y^2_1}{l'^2}.$$

Comparing this with

$$\mathrm{OPD}_p=\tfrac{1}{4}\frac{\mathrm{LA}'_1Y^2_1}{l'^2},$$

the highly important result emerges that at the best focus midway between the paraxial and extreme marginal foci the residual differences of optical paths are of only one-quarter of the magnitude attained by them at either the paraxial or the marginal focus.

The case of simple spherical aberration was one of those on which Lord Rayleigh founded his limit. But he did not extend his work to the paraxial focus and did not observe that there is a point at which the physical aberration sinks to only one-quarter of its paraxial value.

It follows that OPD_p may be allowed to reach a whole wave-length without infringement of the Rayleigh limit for the best focus.

Again, assuming an average wave-length as ·00002″ or ·0005 mm., equation (11), combined with the equation used above, $d(\mathrm{OPD}_p)/dY=4\mathrm{OPD}_p/Y$, may be used to determine the values of the geometrical aberrations at which the Rayleigh limit is reached. We have

$$[\mathrm{AA}']\frac{Y}{4}=\mathrm{TA}'\cdot\frac{Y}{4l'}=\mathrm{LA}'\cdot\frac{Y^2}{4l'^2}$$

$$=\pm\cdot00002''=\pm\cdot0005\text{ mm.}$$

These formulae are most conveniently evaluated for the ratio of focal length to aperture—the f-numbers of photographic optics—represented by $l'/2Y$. As regards TA′ it is well known that the "circle of least confusion" of geometrical optics has a diameter equal to $\tfrac{1}{2}$TA′ and

$$\frac{\mathrm{TA}'}{2}=\mathrm{LA}'\cdot\frac{Y}{2l'}=\frac{\mathrm{LA}'}{4\text{ times the }f\text{ number.}}$$

This is directly comparable with the resolving power, which, in accordance with microscopical experience, is defined by the product of the wave-length and the f-number and means that two points at that distance in the image can just be seen apart.

TABLE I

RAYLEIGH LIMIT FOR SIMPLE SPHERICAL ABERRATION

$\frac{l'}{2Y}=f$-number.	$f/8$.	$f/16$.	$f/32$.
Permissible LA′ .	±·020″	±·082″	±·33″
$\tfrac{1}{2}$TA′ = diameter of geometrical circle of least confusion . .	·00063	·0013	·0026
Actual resolving power = f-number × wave-length .	·00016	·00032	·00064

The permissible residue of longitudinal spherical aberration is seen to be surprisingly large, especially for the slender image-forming cones (rarely wider than $f/16$) of microscope objectives. The geometrical circle of least confusion comes out at exactly four times the resolving power of the system, indicating the erroneous results obtained from the geometrical theory.

The realisation of the full resolving power stated in the last line of the table at the present time rests not only on direct observation and on actual experience in designing lens systems on the basis of the Rayleigh limit, but has also been verified by direct calculation of the light distribution in the plane of the best focus and in the neighbourhood of that plane.

It might at first sight appear that the "tolerance" for primary spherical aberration was of no interest in the discussion of highly corrected instruments because they are expected to be free from this defect. They, however, are so only if used in accordance with the intentions of the designer. In practice large liberties are taken by using microscope objectives at tube-lengths widely departing from that for which they were corrected, and in telescopes both by using eyepieces afflicted with considerable spherical aberration and by using instruments designed for distant objects for the observation of laboratory instruments only a few yards away. In all these cases primary spherical aberration makes its

appearance. Table I. explains why satisfactory results are usually obtained in spite of it.

(ii.) *The Rayleigh Limit for Zonal Aberration.*—The aperture of nearly all lens systems is so large that higher terms of the aberration series become sensible. In the case of low and medium power microscope objectives up to about ·35 NA these higher aberrations cannot be corrected, and it becomes important to limit the aperture to just that value at which the higher aberration does not sensibly lower the resolving power. This ideal can be closely realised by applying the Rayleigh limit.

The usual and, as it happens, the best possible method of designing such systems is so to proportion them that the paraxial and the extreme marginal geometrical rays are brought accurately to the same focus; the trigonometrically determined aberration of the marginal ray is therefore brought to zero. In accordance with equation (11) this means that $d(\mathrm{OPD}_p)/d\mathrm{Y}$ is also zero. As paraxial and marginal foci coincide, OPD_m and OPD_p are identical in this case, hence OPD_p can be directly determined by computing equation (10**) for all surfaces of the system and forming the algebraical sum. If higher aberration is present and predominantly of the Y^6 order, this sum will have a sensible value. A very low value would in any case indicate a high state of correction of the system, but not necessarily absence of higher aberration, which will have to be tested for as described in the following section if deep curvatures occur in the system.

Assuming that aberrations of higher than the sixth order are of unimportant magnitude, the equation of the distorted wave will be

$$\mathrm{OPD}_p = k_1\mathrm{Y}^4 + k_2\mathrm{Y}^6,$$

and its value for the full aperture Y_1 is known by the computation of (10**). But it is also known by (11) (on account of the geometrical spherical correction) that $d(\mathrm{OPD}_p)/d\mathrm{Y}_1$ is zero. Therefore there are two equations,

$$\mathrm{OPD}_{p_1} = k_1\mathrm{Y}_1^4 + k_2\mathrm{Y}_1^6,$$

and $$\frac{d(\mathrm{OPD}_{p_1})}{d\mathrm{Y}_1} = 4k\mathrm{Y}_1^3 + 6k_2\mathrm{Y}_1^5 = 0,$$

which can be solved for k_1 and k_2, and on putting the values found into the general equation for OPD_p, the latter becomes

$$\mathrm{OPD}_p = 3\mathrm{OPD}_{p_1}\left(\frac{\mathrm{Y}}{\mathrm{Y}_1}\right)^4 - 2\mathrm{OPD}_{p_1}\left(\frac{\mathrm{Y}}{\mathrm{Y}_1}\right)^6.$$

When plotted, the curve of the distorted wave takes the form shown in *Fig.* 8 if, as is invariably the case, the marginal OPD_{p1} is assumed to have a positive value. The distorted wave bends away from the ideal wave for the combined paraxial and marginal foci, reaches a maximum distance from it at the full aperture, and would cut the wave again some

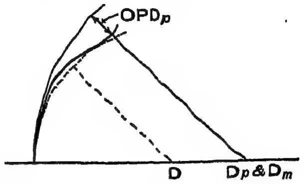


FIG. 8.

distance beyond the full utilised aperture. It is immediately evident that a point D nearer to the lens system must exist which is far more nearly equidistant from all parts of the distorted wave. The closest-fitting ideal wave is dotted in the diagram. It can be determined algebraically by noting that in accordance with equations (10) and (10*) a shift of focus away from D_p will introduce a term in Y^2 into the equation of the wave-curve. Strictly speaking, the terms in Y^4 and Y^6 would be slightly changed at the same time, but if the residual aberrations are as small as they must be if an instrument is to be of practical use, this small change may be safely neglected. A somewhat complicated investigation of the minimum problem involved leads to the result that the residual differences of optical paths become the smallest possible if the point D is so selected (which in the use of the instrument is done automatically and unconsciously simply by searching for the sharpest image) that for point D the residual wave distortion is of the form

$$\mathrm{OPD} = -1\tfrac{1}{8}\mathrm{OPD}_{p_1}\left(\frac{\mathrm{Y}}{\mathrm{Y}_1}\right)^2 + 3\mathrm{OPD}_{p_1}\left(\frac{\mathrm{Y}}{\mathrm{Y}_1}\right)^4 - 2\mathrm{OPD}_{p_1}\left(\frac{\mathrm{Y}}{\mathrm{Y}_1}\right)^6.$$

This gives the same highest value $= -\tfrac{1}{8}\mathrm{OPD}_{p1}$ for half and for full aperture and zero-value in the axis and at $\sqrt{\tfrac{3}{4}}$ of the full aperture. The residual phase-differences in the case of zonal aberration at the best focus are therefore only one-eighth of the amount found by (10**) for the marginal ray. As, in accordance with the Rayleigh limit, a difference of one-quarter wave-length may be allowed, it follows that in lens systems suffering from ordinary zonal aberration (that is Y^6 aberration in the absence of sensible amounts of still higher aberration) the OPD_m calculated by (10**) may be allowed to reach two whole wave-lengths. This result also has been tested and confirmed by a large number of successful lens designs based upon it and more recently by direct integration for the complete light distribution at the best focus.

In the more usual purely trigonometrical method of designing lens systems the zonal aberration is searched for by tracing a third ray through $\sqrt{\tfrac{1}{2}}$ of the full aperture in addition to the paraxial and extreme marginal rays. As is easily shown, the zonal longitudinal spherical aberration reaches a maximum at $Y = \cdot 7071\, Y_1$ if aberrations higher than the Y^6 term are absent. The amount of this zonal longitudinal aberration corresponding to the Rayleigh limit can be investigated in a manner similar to that employed above for ordinary spherical aberration. The result arrived at is that the longitudinal aberration for the ·7071 ray may reach 1½ times the amounts of the LA′ stated in Table I. The corresponding geometrical "circle of least confusion" measures about nine times the resolving power of the system, so that the geometrically estimated size of the image is even more misleading in this case than in that of ordinary spherical aberration.

The analytical method of removing the fourth order aberration completely leads to a wave-distortion of the simple form

$$OPD_p = k_2 Y^6,$$

k_2 having very nearly the same value as it would have if the trigonometrical method were adopted for the same system. A better focus can be found in this case at some distance beyond the paraxial one, but the residual differences of phase at this best possible focus are found to be 6·16 times as large as those existing at the best focus of a trigonometrically corrected system. As the zonal aberration grows with the 6th power of the aperture this means that a trigonometrically corrected lens system may be given $\sqrt[6]{6 \cdot 16} = 1 \cdot 35$ times the aperture possible with an analytically corrected system of the same type.

(iii.) *The Rayleigh Limit for Higher Zonal Aberration.*—In microscope objectives of NA exceeding about ·4 the Y^6 aberration can be controlled. Its negative value in objectives of low NA is due to the excessive marginal over-correction produced by the dispersive contact surfaces. This effect is always present, but when the NA is large and the usual plano-convex front lens is adopted, then very large angles of incidence accompanied by heavy spherical under-correction occur at the first plane surface of the system. The positive Y^6 aberration produced at this surface can then be played out against the negative Y^6 aberration of the corrective contacts, and by securing a suitable free working distance any desired balance may be struck. In the purely trigonometrical method of computing objectives it is usual to aim at union in one point of the rays passing through the paraxial, the marginal, and the ·7071 zone of the full aperture. It is, however, both simpler and leads to a closer knowledge of the residual aberrations to calculate a paraxial and a marginal ray and to bring these to a common focus and then to aim at making the OPD = zero for the computed marginal ray. If only Y^4 and Y^6 aberration were present the correction would then be perfect. In reality the higher aberrations come in in rapid succession. A balance can then be established between the Y^4 and Y^6 aberrations which are under control and the still higher aberrations which cannot be controlled. A full discussion is only possible by making the assumption that Y^8 aberration is the only one of sensible magnitude. This certainly is only a rough approximation, and for that reason the discussion must not be taken as absolutely trustworthy.

It is assumed that the wave-distortion is of the form $OPD_p = k_1 Y^4 + k_2 Y^6 + k_3 Y^8$. The correction stipulated is that the OPD_{p1} for the computed marginal ray shall be zero, and as the geometrical aberration is also removed for the same ray $d(OPD_{p1})/dY_1$ will also be zero in accordance with equation (11). This gives the two equations,

$$k_1 Y_1^4 + k_2 Y_1^6 + k_3 Y_1^8 = 0,$$

and

$$\frac{d(OPD_{p1})}{dY_1} = 4k_1 Y_1^3 + 6k_2 Y_1^5 + 8k_3 Y_1^7 = 0,$$

by which k_1 and k_2 can be expressed in terms of k_3, namely

$$k_1 = k_3 Y_1^4, \quad \text{and } k_2 = k_3 Y_1^2.$$

On putting this into the general expression for OPD_p it becomes

$$OPD_p = k_3 Y_1^4 Y^4 - 2k_3 Y_1^2 Y^6 + k_3 Y^8,$$

and this is zero for $Y = 0$ and for $Y = Y_1$, and has a maximum value for $Y = \cdot 7071\, Y_1$ and a minimum for $Y = Y_1$. For the latter the OPD is, of course, zero. For the maximum at $Y = \cdot 7071\, Y_1$ the value is

$$OPD_{p\,\text{maximum}} = \tfrac{1}{16} k_3 Y_1^8.$$

The residual greatest phase-difference at the computed focus is therefore $\tfrac{1}{16}$ of the total Y^8 aberration present in the system. The curve representing this wave is shown in *Fig.* 9.

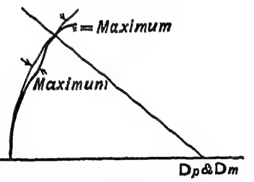


FIG. 9.

No improvement can be made in this case by a shift of focus. But as the zero-value of the OPD at the computed aperture is a minimum, the wave-curve bends again away from the ideal wave beyond the computed aperture, and the aperture may therefore be increased up to the point where the OPD reaches the same value as at ·7071 Y_1 without any increase in the maximum phase-differences at D_p. It is easily ascertained that the general expression for OPD_p reaches the value $\tfrac{1}{16} k_3 Y_1^8$, again at $Y = 1 \cdot 10\, Y_1$. Systems in which

this highest correction is aimed at should, the efore, have the "marginal" ray traced through $\frac{11}{10}$ of the intended *full* aperture. The advantage secured in this way is far greater than it would appear to be at first sight, for at $Y=1\cdot10\ Y_1$ the Y^8 aberration, which is the one beyond control by which the attainable NA is limited, will be $(1\cdot10)^8 = 2\cdot144$ times as large as at $Y=1\cdot00$. By establishing the stipulated corrections for $\frac{11}{10}$ of the intended full aperture, the fraction of the Y^8 aberration present in the real marginal zone which becomes effective at the focus therefore sinks to $1/(16\times2\cdot144)$ or to $\frac{1}{34}$ part. The Y^8 aberration at the extreme margin might therefore reach $\frac{1}{4}$ of 34 or $8\frac{1}{2}$ wave-lengths without transgression of the Rayleigh quarter-wave limit.

Whilst a highly favourable balance of the higher aberrations is sure to be established by the above method, no clue is afforded by it as to the amount of Y^8 aberration which may be present. This may, however, be closely estimated by noting that at aperture Y_1 the Y^4 aberration is exactly equal to the Y^8 aberration. The Y^4 aberration can be determined with sufficient approximation by reducing equation (10**) to first order terms by putting sines equal to angles and cosines equal to one. It then becomes

$$\text{Paraxial OPD}_m = \tfrac{1}{2}N'yi'(u'-u)(i-u'),$$

and if this is calculated with the nominal paraxial angles stipulated in the trigonometrical part of this article it will by summation over all the surfaces give the required indirect estimate of the Y^8 aberration of the computed ray. As $\frac{1}{16}$ of this appears as wave-distortion at the focus, this paraxial OPD_m sum may be allowed to reach 4 wave-lengths, but, on account of the uncertainties introduced by the presence of still higher aberrations and by the rough method of estimating the Y^8 aberration, it is highly advisable to restrict the paraxial OPD_m to a maximum value of only two wave-lengths. This would leave no reasonable doubt as to fulfilment of the Rayleigh condition.

There is no definite method for gaining control over aberrations higher than Y^6; if one were found, then it is an obvious deduction from the preceding paragraph that paraxial $\text{OPD}_m=0$ would be a convenient and effective additional condition to be fulfilled. It would still call for the tracing of only two rays through the system.

§ (11) Focal Range and Depth of Focus.—A very important application of the Rayleigh limit leads to the determination of the range within which the focal adjustment of an optical instrument may vary without sensible loss of definition and conversely to the fixing of the "depth of focus," that is the distance from the nearest to the farthest object which can be seen sharply at any one focal adjustment. All that is required is to discuss equation (11*)

$$\text{OPD} = \text{OPD}_p - \tfrac{1}{2}df\left(\frac{Y}{l'}\right)^2.$$

on the supposition that OPD_p has a fixed value, and that df is subjected to variation. The discussion is simplest for a perfectly corrected instrument in which OPD_p is zero. The Rayleigh limit of $\frac{1}{4}$ wave-length will then apply to OPD, hence

$$df = \tfrac{1}{2}\text{ wave-length}\left(\frac{l'}{Y}\right)^2.$$

By comparison with the permissible LA′ in the case of simple spherical aberration the value of df is seen to be $\frac{1}{8}$ of that found for LA′. But df may obviously be applied on either side of the geometrical sharp focus, hence for a perfectly corrected instrument

Focal Range $= \frac{1}{4}$ of the LA′ in Table I.

The testing of this result by direct integration brings out the same remarkable fact, which also appears in all the other tested cases of a phase-difference equal to the Rayleigh $\frac{1}{4}$ wave-length, namely, that the small deterioration of the image as compared with Airy's ideal spurious disc is limited to a loss of brightness in the central condensation (which nearly always is the only part appreciated by the eye) but does not affect its effective diameter. Within the Rayleigh limit there is thus no loss of resolving power, but only a moderate loss of brilliancy and contrast in the image. On the average (the figures vary sensibly for different cases) the loss of light from the central condensation at the Rayleigh limit is 20 per cent. This remarkable and most valuable peculiarity is maintained up to about twice the Rayleigh limit: the central condensation still maintains its small diameter, but at the doubled limit sinks to about 45 per cent of the ideal maximum brightness, all the light lost in the central condensation appearing in a faint halo and in the more or less distinct diffraction rings. For this reason all the tolerances deduced above may be doubled in case of extreme necessity, still without serious loss of resolving power on detail possessing sufficient contrast. The focal range becomes quite large at high f/numbers as it grows with their square. Thus in photomicrography the cones converging towards the sensitive plate are hardly ever wider than $f/128$, and $f/400$ may be taken as near the average. For $f/400$ the focal range within the Rayleigh limit is 625 times the figure given in Table I. for the LA′ limit at $f/8$, or 12·5 inches. It is therefore quite unnecessary to focus on ground glass in order to localise the image. Equally good results will be obtained, with a great saving of time and eyesight, by focussing "in the air" with a weak magnifier (strong reading-glass) held at about its focal length from the eye and adjusted by guess for focus in the plane to be occupied by the photo-

graphic plate. The image seen in this way will have many times the brightness of that seen on ground glass, and the detail will not be obscured by the grain of the focussing screen.

When there are sensible differences of phase due to spherical aberration at the best focus, then the focal range is necessarily correspondingly reduced, and this reduction is in fact the chief drawback attaching to residuals of aberration up to the Rayleigh limit. If the aberrational differences of phase can be restricted to half the Rayleigh limit the loss of focal range is unimportant, and if only one-quarter of the Rayleigh limit is used up to cover aberration the loss of focal range is quite insensible.

The comparison of the physically determined light distribution near a focus with that suggested by purely geometrical ray tracing yields remarkable diagrams. The upper part of *Fig.* 10 shows by the two inclined

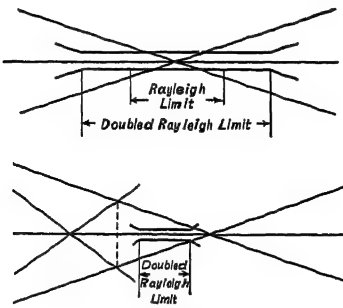


FIG. 10.

straight lines the limits of the geometrical cone of rays in the absence of all aberration. Integration of the interference effect leads to the concentration of all the normally appreciated light between the two parallel lines, which only begin to spread out at the doubled Rayleigh limit. The lower part represents the case of simple spherical aberration at the Rayleigh limit. The inclined lines represent the rays from half and full aperture respectively, the intersection of which determines the geometrical circle of least confusion. The physically determined light distribution is represented by the close parallel lines which lie entirely to the right of the geometrically determined supposed best image and enclose a cylinder of only one-quarter the diameter of the circle of least confusion.

It is noteworthy that Taylor, in his *System of Applied Optics*, called attention to this cylindrical constriction of the light near a focus as an observational fact worthy of theoretical investigation. Any keen observer can easily verify it.

The depth of focus in the object space may be deduced directly from the focal range in the image space by the well-known theorem that the magnification in depth (along the optical axis) is the square of the linear magnification. This, however, becomes inaccurate for high numerical apertures, and it is then desirable to determine df at the object directly by the strict formula from which the above deductions were derived:

$$\text{Difference of phase} = 2df \sin^2 \frac{DD_pD'}{2}.$$

On the side of the object the angle DD_pD' is $\sin^{-1}NA$ if the object is in air, or $\sin^{-1} NA/N$ if the object lies in a medium of index N. In the latter case the wave-length at the object will also be shortened in the proportion of 1 to N, hence the universally applicable equation for df will be

$$df = \frac{\text{Allowed difference of phase}}{2N \sin^2 (\frac{1}{2} \sin^{-1} (NA/N))},$$

in which the difference of phase (according to Rayleigh $= \frac{1}{4}$ wave-length) must be measured by the wave-length in air. The range df may be allowed on either side of the sharply focussed object, hence for the Rayleigh limit

$$\text{Depth of focus} = \frac{\cdot000005 \text{ inch}}{N \sin^2 (\frac{1}{2} \sin^{-1} (NA/N))}.$$

This gives for the numerical apertures stated in the first horizontal line the values in Table II. for wave-length ·00002 in. (bluish green):

TABLE II

Numerical Aperture.	Depth of Focus in Air.	Depth of Focus in Medium of 1·5 Index.
	inch.	inch.
·25	·000312	·000476
·50	·000075	·000117
·75	·000030	·000050
1·00	..	·000026
1·25	..	·000015

The small values of the depth of focus explain the necessity of a delicate, fine adjustment for the focussing of microscopes. The amounts given in the table are available even in the case of projection of the image upon a fixed screen or photographic plate. In visual observations there is an additional amount of depth of focus due to the range of accommodation of the human eye: this is dealt with in the introduction.

§ (12) THE OPTICAL SINE CONDITION.—Although most perfectly freed from spherical aberration, a microscope objective may be utterly useless on account of coma in the images of objects not lying exactly in the optical axis. In microscope objectives of high numerical aperture this defect can reach such extraordinary magnitude that the theoretical resolving power may only be realised if the two close points are placed symmetrically to either side of the optical axis. The removal of coma is therefore

absolutely indispensable. Until 1873 this called for laborious calculations or else for numerous empirical trials of experimental lenses. In that year Abbe and Helmholtz simultaneously but quite independently announced the remarkable theorem known as the optical sine condition, by which the detection of coma is reduced to an almost negligible amount of computation. The theorem states that when any centred lens system causes a ray starting under an angle U from an axial object-point B (*Fig.* 11) to

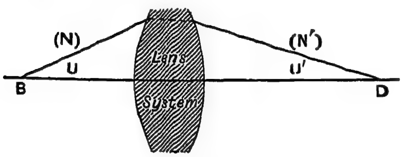


FIG. 11.

reach the trigonometrically determined image-point D under angle U′, then the ratio N sin U/N′ sin U′ will be the magnification of the image produced in the plane of D of a small object near B and in the plane passing through it at right angles with the optical axis. N represents the refractive index of the medium surrounding the object, and N′ the index of the medium in which the image is formed. It follows at once that if zones of the lens system of greater or smaller diameter are considered, they will give an image of the same size only if the ratio N sin U/N′ sin U′ is constant throughout the full aperture. All that is required in order to remove coma in the small field employed by microscope and telescope objectives is therefore to form the ratio N sin U/N′ sin U′ for every computed ray (usually the paraxial and the marginal one), and to modify the system until this ratio attains the same value for the several rays. This is in fact the universal practice of designers of telescope and microscope objectives, and it is very rarely the case that any suspicion of residual coma can be detected in systems thus corrected. When this happens it is due either to zonal variation of the sine ratio or to higher forms of coma which grow with the third power of the diameter of the field and are not included in the theorem. The latter covers all orders of simple coma growing in direct proportion with the diameter of the field.

Many proofs have been given of this remarkable theorem, but most of them are either incomplete or of a highly involved type. For a simple and yet valid proof reference is made to a paper by Conrady in Monthly Notices of *R.A.S.* for March 1905.

§ (13) THE CHROMATIC CORRECTION OF OBJECT GLASSES.—All the ray-tracing equations which have been given involve the refractive index of the lenses of which a lens system is built up. As the refractive index of any given glass varies for light of different colour and wave-length, the conclusion is obvious that the location of the final focal point, the magnification, and the spherical correction will also, as a rule, vary according to the colour of the light. Owing to the comparatively small change of N in the visually bright part of the spectrum (averaging less than 1 per cent of N between the C and F lines) and the pronounced maximum of intensity at the middle of this short range, the variations of magnification and spherical aberration are usually small and can be neglected in systems of low aperture, or treated by a lower approximation than that required for the spherical correction of the brightest rays in systems of high aperture.

The makers of optical glass supply with each melting a table giving the refractive index N_d for the D-line, the dispersion $N_f - N_c = dN$ between the C and F lines, $N_f - N_d$ between D and F, and under the symbol small Greek ν, the ratio of $(N_d - 1)$ to $(N_f - N_c)$. $N'_g - N_f$ for the blue end of the spectrum is also frequently given, G′ being the adopted symbol for the dark-blue hydrogen line close to the solar G-line. Other indices have to be found by interpolation. The best formula (Conrady, Monthly Notices, R.A.S., 1903) for this purpose is

$$N = N_0 + aw + bw^{3\frac{1}{2}},$$

in which N_0 represents the hypothetical index for infinite wave-length, w the reciprocal of the wave-length expressed in μ (·001 mm.), for which N is to be determined, and a and b constants for any one glass. By writing out this formula for the values of N usually given by the glassmakers (C, D, F, and G′), and introducing on the right the corresponding numerical values of w, N_0, a, and b can be solved for in terms of the usual data, and are found as

$$N_0 = N_d - [\cdot 84895](N_f - N_c) + [\cdot 92452](N'_g - N_f),$$

$$a = [\cdot 69108](N_f - N_c) - [\cdot 81775](N'_g - N_f),$$

$$b = -[9\cdot 29980 - 10](N_f - N_c) + [9\cdot 63529 - 10](N'_g - N_f).$$

The figures in square brackets are the logarithms of the numerical factors, which are more useful than the factors themselves.

For the whole range of ordinary optical glasses this interpolation formula gives indices agreeing with direct determinations within one or two units of the fifth decimal place throughout the visible spectrum. On the basis of the nature of the secondary spectrum,

the formula can be discussed so as to determine the wave-length for which a lens system has minimum focal length when the C and F rays are brought to a common focus in accordance with almost universal custom based on long experience. This wave-length of the visually most important rays is thus found as 5555 Ångström units, and the corresponding index can be determined accurately by

$$N_{5555} = N_d + \cdot 2157 (N_f - N_c) - \cdot 0474 (N'_g - N_f),$$

or with nearly always sufficient approximation by the extremely simple formula

$$N_{5555} = N_d + \cdot 188 (N_f - N_c),$$

which is obtained from the more accurate one by introducing ·585 as an average value of the ratio $(N'_g - N_f)/(N_f - N_c)$, from which the latter does not vary for practically useful glasses by more than about ±6 per cent. Decidedly better results are obtained by carrying out the main calculation of lens systems for fulfilment of the sine condition, and freedom from spherical aberration within the Rayleigh limit with this index for visually brightest light obtainable by a single slide-rule setting instead of the widely used N_d.

§ (14) Geometrical Correction of the Chromatic Aberration.—Ray-tracing methods for dealing with the chromatic aberration must be devised with due regard to the variation of spherical aberration in different colours, and to the existence of the secondary spectrum.

A first method consists in tracing a paraxial and a marginal ray in each of two colours near the limits of the spectrum, which is effective under the given conditions. C and F are generally best for visual purposes, D and G′ for systems intended for photography at the visually determined focus. It will nearly always be found that the four rays cannot be brought to a common focus: the best that can be done is to allow spherical over-correction for one (practically always the more refrangible) colour and spherical under-correction of nearly the same amount for the other (less refrangible) colour, and to bring the four intersection points within the smallest possible space along the axis, and to the most symmetrical distribution, which will be usually the sequence, counting in the direction away from the lens system: blue paraxial focus closely followed by the red marginal focus, after a fairly considerable interval the red paraxial focus, and closely beyond the latter the blue marginal focus. As the difference in the spherical aberration for the two colours is almost entirely of the primary order, this arrangement of the foci implies that the red and blue rays through the ·7071 zone would be found closely united at the midway point. The chief drawback of this method of carrying out the chromatic correction is, in the case of microscope objectives, that the criteria for the Rayleigh limit cannot be conveniently applied on account of the residual longitudinal spherical aberration in both colours. Its advantage is that the spherical variation of chromatic aberration is determined directly.

A better method consists in the tracing of a paraxial and a marginal ray in "brightest" light, as defined for visual purposes by the simple interpolation formula given above, and in establishing perfect correction of the longitudinal spherical aberration for these two rays. To correct the chromatic aberration two rays of different colour (C and F for visual purposes) are traced through the ·7071 zone and brought accurately to a common focus. On account of the secondary spectrum the latter will lie at some distance beyond the focus of the brightest light. In this case the criteria for the Rayleigh limit can be applied to the brightest light. The spherical variation of chromatic aberration is, however, not revealed.

A very convenient simplification of the preceding method results if the effect of the secondary spectrum is eliminated from the *calculation* (it of course remains in the actual lens system) by coupling with the calculation of the paraxial and marginal "brightest" rays one fictitious "violet" ray traced through the ·7071 zone, the N_v used being determined for each glass as the sum of its index for brightest light plus its dispersion between C and F. In this method the three computed rays can be brought accurately to one common focus. It can be confidently recommended for all ordinary purposes and for low-power microscope objectives if a purely ray-tracing method is preferred for any reason to the more expeditious and physically sound optical path method now to be given.

§ (15) Physical Treatment of the Chromatic Aberration.—Referring back to *Fig.* 2, but assuming that the object-point O send out white light composed of all colours of the spectrum, the different velocities of propagation of these colours in any one lens of the system, and along axial and marginal paths in it of different length, will in general cause emergence of waves of varying curvature corresponding to the separate colours. The aim of achromatisation is to reduce these variations of curvature to the least possible magnitude for the range of the spectrum which contributes most strongly to the final image. *Fig.* 12, an elaboration of *Fig.* 2, will make this quite clear. If the thick curve W′ in the image-space represents the emerging wave of brightest light, then in the case of chromatic under-correction the emerging more refrangible waves will have a greater curvature—and therefore a shorter intersection length of corresponding rays—and the less refrangible waves will have less curvature. By small adjustments of the relative power of crown and flint components

the marginal gap between corresponding waves of different colour can be modified, and as the differences of refractive index for the range of

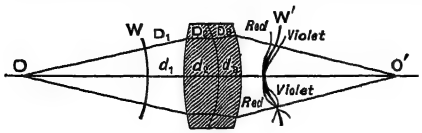


FIG. 12.

colours which calls for correction is always small, the *relative* distortion of the waves will be only affected to an unimportant extent by such a change. Under these conditions reasoning of the same kind as was applied above in the application of the Rayleigh limit to the problems of spherical aberration and the location of the best focus in those cases immediately leads to the conclusion that the best compromise (*i.e.* the least residual gaps between the different waves in the intermediate zones) will be attained if the waves of different colour are so adjusted that they intersect each other in the extreme margin of the effective aperture when they are tangent to each other in the axis. This leads immediately to a solution of the problem. By the definition of a wave surface the optical paths along the geometrical rays between conjugate points of two positions of a wave are absolutely equal. If the upper ray in *Fig.* 12 represents the trigonometrically traced marginal ray in brightest light, and d_1, d_2, etc., are successive parts of the axial path, D_1, D_2, etc., the corresponding parts of the marginal path, and if N_1, N_2, etc., are the respective refractive indices for brightest light, the relation therefore exists for the two paths between W and W′

$$\Sigma(d-D)N=0.$$

If the lower marginal ray in the diagram is similarly regarded as the trigonometrical tracing from W to the wave tangent to W′ in the axis, but of a different colour, and if D'_1, D'_2, etc., are the successive parts of the marginal path, and N_1+dN_1, N_2+dN_2, etc., the respective indices, then for this coloured wave the corresponding relation exists

$$\Sigma(d-D')(N+dN)=0.$$

The discussion of the two sums would still call for the trigonometrical tracing of the two rays of different colour on account of the difference between the corresponding D and D′. At this stage Fermat's "theorem of the minimum optical path," as it is loosely called, becomes available. According to this theorem, the optical path of light between two points as determined by the law of refraction, that is by exact trigonometrical ray-tracing, represents a maximum, minimum, or stationary value with reference to all alternative closely neighbouring paths. These neighbouring paths therefore can at most differ by small quantities of the second order if they are everywhere within small distances and at small angles of the first order compared to the computed path. It follows that if dN represents a very small increment of N, then the differences between the corresponding sums of D and D′ values are strictly negligible. But the same convenient simplification is found to hold for any finite value of dN which can occur in lens systems. When the difference between a trigonometrically determined path and a neighbouring path at small but finite distances and angles is evaluated it is easily proved that this difference will rarely amount to so much as $\frac{1}{10}$ wave-length for any thicknesses or any angles of incidence that can occur in practicable lens systems. The neglecting of the difference between each D and the corresponding D′ for each constituent lens or space is therefore always legitimate in actual lens systems.

For the ideal chromatic correction shown in the lower part of *Fig.* 12, when the coloured and the brightest wave intersect in the marginal zone this equality of D and D′ applies to all sections of the total path. There are then the two relations—

For the brightest light, $\Sigma(d-D)N \quad =0,$

For the coloured light, $\Sigma(d-D)(N+dN)=0,$

and the difference, $\Sigma(d-D)dN \quad =0,$

embodies the condition to be fulfilled by the lens system if the ideal chromatic correction is to be realised.

If the ideal correction shown in the lower part of *Fig.* 12 is not realised there will be a marginal gap between the brightest and the coloured wave, and for the last section of the respective marginal paths there will then be a corresponding difference between D and D′. If the sum of the $(d-D)dN$ is still calculated with the D-values for the brightest light, the sum will differ from zero by the exact amount of the marginal gap between the two waves. Therefore this sum is an unconditional measure of the residual chromatic aberration of any lens-system if put into the form

$$\Sigma(d-D)dN=\text{Chromatic aberration}, \quad (12)$$

the latter expressed as a difference of optical paths.

The sense of the difference $(d-D)$ has been chosen so as to give a positive value for a simple convex lens $(d>D)$ and for coloured light more refrangible than the brightest light $(dN>0)$ in accordance with the custom of counting chromatic "under-correction" as positive in this case.

The indices and dispersions supplied by the glassmakers are determined in air, and are therefore relative and not absolute values. This fact justifies to a large extent the universal custom in practical optics of computing lens-systems with a total disregard of the refractive properties of air. The air is always treated as if its refractive index were exactly *one* and its dispersion *zero*. In evaluating equation (12) this means that air-spaces contribute nothing at all to the value of the chromatic sum, and that the terms $(d-\mathrm{D})d\mathrm{N}$ have only to be calculated for the actual simple lenses of which the system is composed. When the space between the object or image and the nearest lens is filled with a medium other than air, as, for instance, with oil or water in the case of immersion objectives and condensers for the microscope, this space must of course be included in forming the sum.

The only quantity in (12) which calls for an addition to the computing formulae is D. By projecting the D of any one lens upon the optical axis it is easily verified that if X_0 is the depth of curvature calculated by (8) of the first surface, X that of the second surface, and U the angle of convergence of the marginal ray, then

$$\mathrm{D}=(d+\mathrm{X}-\mathrm{X}_0)\sec \mathrm{U}. \qquad (12^*)$$

This formula also covers the case of the space between the object-point and the first lens surface of an immersion-system, by looking upon this space as a lens consisting of immersion-medium and of a thickness equal to the distance from the object-point to the pole of the first actual lens surface. X_0 is then obviously zero and the formula becomes simplified to that extent.

A well-handled slide-rule is generally sufficient for the working out of the $(d-\mathrm{D})d\mathrm{N}$ sum, because the $d\mathrm{N}$ supplied with optical glass are neither determined with sufficient accuracy nor sufficiently constant for different plates of the same melting to be depended upon beyond about 1 part in 500. Unusually thick lenses will, however, call for closer calculation, as in their case the percentage uncertainty of $(d-\mathrm{D})$ will be much higher than that of D. Even then the slide-rule may be used, provided that U is reasonably small, by calculating directly

$$d-\mathrm{D}=(\mathrm{X}_0-\mathrm{X})\sec \mathrm{U}-d(\sec \mathrm{U}-1),$$

in which the great thickness appears with a very small factor and influences merely a small correction.

In practice the $(d-\mathrm{D})d\mathrm{N}$ is calculated for visual instruments with $d\mathrm{N}=\mathrm{N}_f-\mathrm{N}_c$ as given by the makers and for photographic systems with $d\mathrm{N}=\mathrm{N}'_g-\mathrm{N}_d$.

§ (16) The Rayleigh Limit for Chromatic Aberration.—As by (12) the $(d-\mathrm{D})d\mathrm{N}$ sum expresses directly the difference of marginal optical paths for the directly traced "brightest" ray and for the coloured ray of a different refractive index, it is at once apparent that the sum may be allowed to differ from zero-value by the Rayleigh quarter-wave provided that the $d\mathrm{N}$ used refers to the colour, most remote from the brightest, which still contributes a sensible amount to the total intensity of the image. In visual instruments the light corresponding to the C and F lines may be taken as representing this limit. But as the practice is to use the whole dispersion between C and F as the value of $d\mathrm{N}$, whilst the visually brightest light corresponds to the region about midway between C and F, a value of the $(d-\mathrm{D})(\mathrm{N}_f-\mathrm{N}_c)$ sum of half a wavelength will correspond to the Rayleigh limit for either C or F light at the focus of the brightest light. For visual instruments of moderate aperture, in which the spherical variation of chromatic aberration may be assumed to be unimportant, the Rayleigh limit therefore corresponds to

$$\Sigma(d-\mathrm{D})(\mathrm{N}_f-\mathrm{N}_c)=\pm\cdot 00001''=\pm\cdot 00025\text{ mm.}$$

As in the case of the spherical tolerances previously discussed, this apparently minute latitude in the value of the chromatic sum really amounts to an extremely generous allowance. When it has to be drawn upon in order to improve the correction of other aberrations in systems of few surfaces (and it should, of course, be taken advantage of only for some good and sufficient reason!) it is usually used up in *one* of the constituent lenses. Now $(d-\mathrm{D})$ may be taken roughly as of the order of ·1 inch for any one average lens and $(\mathrm{N}_f-\mathrm{N}_c)$ has an average value of ·01. The full amount of $(d-\mathrm{D})(\mathrm{N}_f-\mathrm{N}_c)$ for an average lens is therefore of the order of ·001 inch. The tolerance of $\pm\cdot 00001$ inch represents 1 per cent of this value, so that either the power of the lens—which is proportional to $(d-\mathrm{D})$—or the dispersion of the glass may be varied to the extent of ± 1 per cent without transgression of the Rayleigh limit, and therefore without sensible loss of defining or resolving power.

In systems of large aperture, which almost invariably contain surfaces of deep curvature, the spherical variation of the chromatic aberration represented by the separation of the different coloured waves in the lower part of *Fig.* 12 for zones between axis and margin must be estimated and restricted within safe limits.

Owing to the smallness of the variation of the refractive index for the range of colours which call for serious consideration in any one instance (C to F for visual instruments), the gap in question is almost entirely due to variation of *primary* spherical aberration for different colours and can be evaluated with reference to the "brightest" wave on the principles which were employed in the section on spherical aberration for the determination of the gap between a wave distorted by

spherical aberration and the closest-fitting ideal spherical wave.

As the brightest and the coloured wave intersect each other in the marginal zone, and as the gap between them is attributable (with sufficient approximation) to primary spherical aberration, the gap must correspond to an equation of the form

$$\text{Gap} = c\left(\frac{Y}{Y_1}\right)^4 - c\left(\frac{Y}{Y_1}\right)^2,$$

which gives a maximum value of $-\frac{1}{4}c$ at $Y = Y_1\sqrt{\frac{1}{2}}$, where Y_1 is used as a symbol for the full semi-aperture.

In the paraxial region only the second term will be sensible, as Y/Y_1 will be a very small fraction. Hence a determination of the $(d-D)dN$ sum for the paraxial region will determine the second term and thereby the value of c. This calls for the determination of the net thickness $(d-D)$ of each lens in the vicinity of the optical axis. This is easily derived from the equation already used,

$$d - D = (X_0 - X)\sec U - d(\sec U - 1).$$

$X_0 - X$ will be very small and the factor sec U in the *first* term is therefore to be neglected, as the term is clearly small of the second order. But in the second term d represents a constant finite quantity multiplied by (sec U − 1), which for small angles is small of the second order. Both terms of the equation, therefore, contribute terms of the second order. For the paraxial value of X equation (8*p*) gives

$$x = \tfrac{1}{2}y(u+i),$$

or, as by *Fig.* 1,

$$y = r(u+i),$$

$$\therefore\ x = \tfrac{1}{2}r(u+i)^2.$$

For small angles (sec U − 1) becomes $= \frac{1}{2}u^2$, hence the computing formula is

$$(d-D)_{\text{paraxial}} = \tfrac{1}{2}r_0(u_0+i_0)^2 - \tfrac{1}{2}r(u+i)^2 - \tfrac{1}{2}d\,.\,u^2. \qquad (12p)$$

If the zonal chromatic aberration is to be determined this equation has to be evaluated for each constituent lens, just like the marginal $(d-D)$, and the sum $(N_f - N_c)(d-D)_{\text{paraxial}}$ formed. The use in this calculation of the large fictitious values of the paraxial angles recommended in the first part of this article is equivalent to determining the Y^2 term of the equation for the gap at full aperture; hence the paraxial $(d-D)dN$ sum gives the value of c in the equation for the gap directly without any further reduction. As the maximum width of the gap has been determined above as equal to $\frac{1}{4}c$, one-quarter of the paraxial $(d-D)dN$ sum represents the maximum distance between the red and blue emergent waves, both of which will be at approximately half this distance from the "brightest" wave between them. In accordance with the Rayleigh limit the paraxial $(d-D)dN$ sum may therefore be allowed to reach $\frac{8}{4}$ or two entire wave-lengths. The rule to be followed in these cases of zonal variation of the chromatic correction is, therefore:

(*a*) Bring the $(d-D)dN$ sum to zero-value for the marginal ray by suitable changes of radii, thicknesses, or separations, or by selection of glass of appropriate dispersion.

(*b*) Calculate the $(d-D)dN$ sum for the paraxial region with the aid of (12*p*). If it does not exceed two wave-lengths (·00004 inch), the zonal chromatic aberration will be within the Rayleigh limit and practically insensible. Objectives which fulfil this or an equivalent geometrical condition are now usually described as semi-apochromatic.

§ (17) The Secondary Spectrum.—The $(d-D)dN$ method of dealing with the chromatic aberration supplies by far the simplest and clearest means of demonstrating and evaluating the so-called secondary spectrum of all lens systems made from the ordinary optical glasses.

As the $(d-D)$ value of the separate components of a system may by Fermat's theorem be treated as practically constant for the whole visible spectrum, it is at once seen that if $(d-D)dN$ has been made zero for two selected colours, it would also be zero for other colours if the dN-values for other colours of all the varieties of glass used in the system were in a fixed ratio to the respective values of dN for the originally selected colours. There are no glasses suitable for use in microscope objectives which fulfil this condition. As an example, two glasses largely used in microscope objectives may be taken from Chance's list—

No. 6493	$N_d = 1{\cdot}5160$		$N_f - N_c = {\cdot}00809,$
No. 337	$1{\cdot}6469$		${\cdot}01917,$
	$N_f - N_d = {\cdot}00567$		$N'_g - N_f = {\cdot}00454,$
	${\cdot}01376$		${\cdot}01170,$

or ratio of dispersions

2·37, 2·43, 2·58.

It is seen that the ratio of flint to crown dispersions increases greatly towards the violet end of the spectrum. To estimate the resulting confusion of phase for D and G′ when F and C have been brought to a common focus, it is best to calculate from the ratio in the F − C region what the dispersion of the crown-glass ought to be to establish proportionality. The result is that to remove the secondary spectrum the crown ought to have

$N_f - N_c = {\cdot}00809 \quad N_f - N_d = {\cdot}00581$
$N'_g - N_f = {\cdot}00494,$

or an increase of

·00000, ·00014, ·00040.

These differences, multiplied by the $(d-D)$ value of the crown, give directly the secondary spectrum effect as a difference of optical paths. Now it may be taken that the total of $(d-D)$ values for the crown components of an average microscope objective is about ·15 inch. Therefore the secondary spectrum effect in terms of optical path is, referred to the combined F and C focus, for the D-line $\cdot00014 \times \cdot15 = \cdot00002'' = 1$ wave-length, for the G′-line $\cdot00040 \times \cdot15 = \cdot00006'' = 3$ wave-lengths.

In reality one observes close to the focus of the brightest light which is nearly enough (for the present purpose) at D. Referred to this the F and C rays will arrive with a marginal difference of phase of one wave-length and the G′ rays with one of four wave-lengths or respectively four and sixteen times the Rayleigh limit. This demonstrates the extreme seriousness of the secondary spectrum in cases when a long range of colours comes into action and the difficulty of obtaining good photographs with "achromatic" microscope objectives. Practically no advantage results from substituting other ordinary optical glasses for the two selected above. But very different conditions are found when the mineral fluorite is substituted for the usual crown-glass and dense crown or very light flint-glass for the usual dense flint.

Fluorite :	$N_d = 1{\cdot}4338$	$N_f - N_c = \cdot00454$,
Telescope-flint :	$1{\cdot}5237$	$\cdot01003$,
	$N_f - N_d = \cdot00321$	$N'_g - N_f = \cdot00256$,
	$\cdot00708$	$\cdot00575$.

Ratio of dispersions : 2·21, 2·205, 2·245.

Calculating what the fluorite dispersions ought to be for perfect proportionality, the result is

$$N_f - N_c = \cdot00454 \quad N_f - N_d = \cdot00321 \quad \cdot00260,$$

or an increase of

$$\cdot00000, \quad \cdot00000, \quad \cdot00004.$$

The total $(d-D)$ value of fluorite in an apochromatic objective would, however, be larger than for an achromatic objective and may be put at ·25 inch, which leads to practically perfect achromatism for the entire C to F region and a secondary spectrum effect of $\cdot00004 \times \cdot25 = \cdot00001'' =$ half a wave-length for G′. Even G′ is therefore at the doubled Rayleigh limit which still gives a decidedly small and reasonably bright image of a point. This shows that a very great improvement can be effected by the proper use of fluorite in microscope objectives. But it must be borne in mind that the advantages will only be realised if the correction of the zonal spherical and chromatic aberrations is of a correspondingly high order. Clearly realising this, Abbe from the very first coupled removal of these zonal aberrations, which can reach astounding magnitude in microscope objectives, with the demand for removal of the secondary spectrum, and reserved the name Apochromat for systems which fulfil all three conditions. Objectives improperly called apochromatic are quite frequently met with, which when used with nearly their full aperture are easily beaten in visual observations and even in photographic work by carefully computed and made true semi-apochromatics of ordinary optical glass.

§ (18) THE CHROMATIC VARIATION OF MAGNIFICATION.—At the present time all microscope objectives of less than ⅛ inch equivalent focal length have an unachromatic simple front lens followed by chromatically over-corrected back-combinations. *Fig.* 13 represents the simplest possible (and for low powers quite a useful) representative of this type.

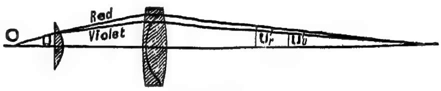


FIG. 13.

A white ray starting from the axial object-point O under angle U will evidently be spectrally decomposed by the unachromatic front lens, and as a consequence the more refrangible components will reach the chromatically over-corrected back lens at points closer to the optical axis than the points at which the less refrangible reach the back lens. If the system as a whole is achromatic for the image-point O′, all the component colours of the original white ray will be brought together at O′, but will necessarily reach it under angles steadily diminishing with the refrangibility of the colours. By the sine condition the magnification produced at O′ will be sin U/sin U′, and will be largest for the most refrangible, smallest for the least refrangible rays, with the obvious result that extra-axial object-points will be rendered not as white points but as short linear spectra with the violet end farthest from the optical axis. If the front lens consists of glass of high dispersion and is widely separated from the corrective back-combination, this chromatic difference of the magnification may reach several per cent of the mean amount and will emphatically call for correction if it approaches or exceeds 1 per cent. It *can* be corrected in the objective itself, as shown in *Fig.* 14, by the addition of a final widely separated uncorrected or chromatically under-corrected lens so calculated that by the action of a heavily over-corrected middle combination the dispersed colours are caused to meet in the final lens and leave it as a reconstituted white ray

proceeding towards O′. Abbe designed some experimental systems on this principle, but it is now invariably preferred to effect the

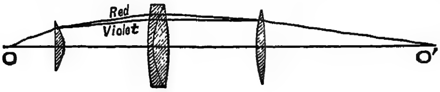


FIG. 14.

correction in the eyepiece by giving to the latter chromatic difference of magnification equal in magnitude but opposite in sign to that of objectives constructed according to *Fig*. 13.

When the chromatic difference of magnification of an objective is to be accurately determined, as it should be in the case of apochromatic systems, the best method consists in tracing a coloured ray right through at $\sqrt{\frac{1}{2}}$ of the full aperture used for the principal calculation.

§ (19) THE ABERRATIONS OF OBLIQUE PENCILS.—A microscope objective spherically and chromatically corrected in accordance with the preceding sections and fulfilling the sine condition will nearly always give satisfactory results throughout the field of small angular extent which is normally utilised. Nevertheless the outer part of the field of low-power eyepieces frequently shows unpleasant indications of the curvature of field and of the astigmatism, which are two defects of extra-axial image-points incapable of full correction in microscope objectives, and which are, partly for that reason, but chiefly on account of the heavy labour involved in their determination, usually left entirely unconsidered in the design of such objectives. Astigmatism causes extra-axial image-points to degenerate into an imperfect concentration of the rays which is characterised by the existence of two focal lines instead of one sharp focal point, one of these focal lines lying in a tangential direction with reference to the centre of the field, and being therefore called the tangential focal line, whilst the other lies at a distance from the first depending on the magnitude of the astigmatism, is directed towards the centre of the field, and is therefore called the sagittal focal line. If the astigmatism is "pure," then the focal lines are perfectly sharp straight lines, and all the rays pass through a small circular disc midway between the two focal lines, whilst at all other points the cross-section of the complete pencil of rays is elliptical. In many cases the astigmatism can be removed by suitable constructional modifications of the lens system. Another defect then remains (with rare exceptions), namely, curvature of the field. This curvature of the field in the absence of astigmatism is determined by the remarkable Petzval theorem. In its original, and hitherto the only generally recognised, form this theorem states that if a centred lens system is free from spherical aberration, coma, and astigmatism, then the residual curvature of the image of a radius $R'i$ is determined by

$$\frac{1}{R'_i}=\frac{1}{R_i}-\Sigma\frac{N-1}{N}\left(\frac{1}{r_1}-\frac{1}{r_2}\right)=\frac{1}{R_i}-\Sigma\frac{1}{N\,.f_0},$$

in which R_i stands for the radius of the surface on which the original object-points lie. The sums have to be taken for all the constituent simple lenses of which the system is built up, N being the refractive index, r_1 and r_2 the first and second radius of curvature, and f_0 the focal length of an infinitely thin lens having these radii. The remarkable feature of the theorem is that the curvature of the image is proved to be totally independent of the thickness and separation of the constituent lenses and also independent of the conjugate distances of object and image.

It has recently been pointed out (Conrady, *Monthly Notices, R.A.S.*, Nov. 1918, Nov. 1919, and Jan. 1920), as an almost obvious deduction from the usual equations for the oblique aberrations of centred lens systems, that the unsatisfactory "if" in the Petzval theorem may be avoided and the utility of the theorem may be greatly increased by incorporating in its statement the fact that in the presence of astigmatism the two focal lines always lie on the same side of the curved image surface defined by the original theorem, and that the tangential focal line of any one oblique pencil always lies at three times the distance of the corresponding sagittal focal line from the Petzval surface. This extended Petzval theorem takes the form

$$\frac{3}{R'_s}=\frac{1}{R'_t}+\frac{2}{R_i}-2\Sigma\frac{N-1}{N}\left(\frac{1}{r_1}-\frac{1}{r_2}\right),$$

in which R'_s signifies the radius of the surface on which all the sagittal focal lines lie, R'_t the corresponding tangential radius, whilst the remaining terms have the significance already referred to. An exact (first approximation) equation of unconditional validity thus takes the place of the widely misunderstood and misinterpreted Petzval theorem. The chief practical value of the extended theorem arises from the fact that it enables a designer to deduce a close value of R'_s as soon as the more easily determined R'_t and R'_i have been found.

In microscope objectives R'_t always has a value nearly equal to the equivalent focal length. If the astigmatism is corrected there will then be a severely curved field, whilst the attempt to reduce this curvature by over-corrected astigmatism (that is, by throwing the focal lines beyond the convex side of the

Petzval surface) leads to loss of definition in the outer part of the forcibly flattened field. The fact that a really satisfactory state of correction is thus impossible justifies to a large extent the universal neglect of a detailed study of the curvature of field and astigmatism in microscope objectives. It is, however, a fact that in the majority of cases, and especially in the spherically and chromatically most highly perfected apochromatic and semi-apochromatic objectives, the already great Petzval curvature is aggravated by under-corrected astigmatism which makes R'_s, and especially R'_t, shorter than R'_p.

The Lister type of low-power objective is decidedly favourable in this respect and can be easily rendered still better by closer approximation to the type of photographic portrait lenses, and this has been done repeatedly, especially for photomicrographic purposes, even to the extent of adopting the anastigmat type of photographic lens. In the higher powers the problem is far more difficult, because the distribution of curvatures of surfaces which favours a flat field is absolutely opposed to that which leads to low residuals of zonal spherical and chromatic aberration in the central part of the field. In the papers already quoted other extensions of the theory of oblique pencils are dealt with which open up new possibilities of reconciling these hitherto contradictory desiderata; but an addition of at least one rather widely separated component to the already complicated high-power objective will nearly always be required and will carry with it a reduction of the transmitted light by about 12 per cent.

Theoretically there is one more defect of lens systems not yet mentioned, namely, distortion: owing to the small angle of field of microscope objectives this defect never reaches appreciable magnitude (except perhaps in the case of very accurate measurements by screw-micrometers), and may therefore be passed over.

§ (20) The Calculation of Microscope Objectives.—It is always preferable to trace the light through a microscope objective in the reverse of the actual direction, that is, from what will eventually be the image to the object. One reason for this procedure is that the tube-length thus becomes a fixed initial datum and will remain unchanged in modifying the system so as to attain the correction of the aberrations. A second justification is supplied by the fact, of which a designer soon becomes aware, that it is far easier and simpler to reach perfect correction by modification of the front lens or lenses than by changes of the compounded back lenses.

All existing microscope objectives, with the possible exception of simple thin cemented achromatic lenses for low magnifications, may be taken to be the result of a succession of purely empirical trials either in actual glass and brass or by the trigonometrical computing method described in an earlier section. For the high powers this method is likely to remain the only one, for it is fairly obvious that in systems in which angles up to 60° or 70° occur no analytical approximation can be of the slightest use. But the lower powers of the Lister type can now be arrived at by a strictly systematic analytical solution recently developed at the Imperial College and based on the theory of oblique pencils given by Conrady in *Monthly Notices of the R.A.S.* of November 1918, November 1919, and January 1920. As an example of a system designed by this process, and also as an illustration of the practical use of the computing formulae given in this article, the following objective of Lister type is selected (*Fig.* 15). The problem set was to design such a system so that it should (in

Fig. 15.

the really employed reverse direction) produce a magnification of −4 times, that the image should be formed at 25 cm. from the back lens, that the separation of the two lenses should be 6·25 cm., and that $\frac{3}{5}$ of the total refraction effect or deviation of the marginal rays should be produced by the back lens. The glasses to be employed were

Chance No. 605	$N_d = 1{\cdot}5175$	$V = 60{\cdot}5.$
Chance No. 360	$1{\cdot}6225$	$36{\cdot}0.$

The analytical solution for freedom from spherical aberration and coma gave for infinitely thin lenses

$$r_1 = 10{\cdot}64 \qquad r_4 = 13{\cdot}51$$
$$(d_1' = {\cdot}4) \qquad (d_4' = {\cdot}3)$$
$$r_2 = 2{\cdot}088 \qquad r_5 = 1{\cdot}449$$
$$(d_2' = {\cdot}2) \qquad (d_5' = {\cdot}2)$$
$$r_3 = 5{\cdot}13 \qquad (r_6 = -3{\cdot}20)$$

Analytically the astigmatism was found to be zero.

This first solution was tested trigonometrically, after putting in the suitable thicknesses given in brackets, by tracing a paraxial and a marginal ray through with initial $l_1 = L_1 = -25$ cm. and $\log u_1 = \log \sin U_1 = 8{\cdot}50504_n$ $(U = -1° 50')$, and gave after passage through the back lens

$$L_3' = 12{\cdot}7020 \qquad U_3' = 3° 37' 0''$$
$$l_3' = 12{\cdot}5096 \qquad u_3' = {\cdot}064120,$$

and after passage through the whole system

$$L_6' = 3{\cdot}0449 \qquad U_6' = 7° 12' 34''$$
$$l_6' = 3{\cdot}0263 \qquad u_6' = {\cdot}125346.$$

As might have been expected, the system proved spherically over-corrected, to the extent of $l_6' - L_6' = -\cdot 0186$. This was corrected by shortening the last radius to $r_6 = -2\cdot 93$ by a few trigonometrical trials. The result then was

$$L_6' = 2\cdot 8696 \qquad U_6' = 7^\circ\ 38'\ 12''$$
$$l_6' = 2\cdot 8699 \qquad u_6' = \cdot 132183.$$

The sine condition then showed the coma correction to be also defective, the magnification produced by the marginal rays being

$$\frac{\sin U_6'}{\sin U_1} = -4\cdot 1539 \text{ times,}$$

that by paraxial rays

$$\frac{u_6'}{u_1} = -4\cdot 1317 \text{ times,}$$

showing an inadmissible difference of about ½ per cent. The analytical solution was therefore repeated by solving for those forms of the two components which would give the opposite of the coma defect found trigonometrically, and the new solution was

$$r_1 = \quad 17\cdot 24 \qquad r_4 = \quad 6\cdot 94$$
$$r_2 = -\ 1\cdot 942 \qquad r_5 = -1\cdot 613$$
$$r_3 = -\ 4\cdot 33 \qquad (r_6 = -4\cdot 13),$$

changed empirically to $-3\cdot 00$.

With the same thicknesses of the individual lenses as before, and the same initial data, this gave

$$L_6' = 2\cdot 9877 \qquad U_6' = 7^\circ\ 14'\ 6''$$
$$l_6' = 2\cdot 9682 \qquad u_6' = \cdot 126509,$$

or, again, spherical over-correction. By two trials the last radius was brought to $r_6 = -3\cdot 60$, and this gave

$$L_6' = 2\cdot 7847 \qquad U_6' = 7^\circ - 44' - 52''$$
$$l_6' = 2\cdot 7840 \qquad u_6' = \cdot 134885.$$

The slight spherical over-correction may safely be ignored as it amounts to only a small fraction of the "tolerance" deduced in an earlier section.

The sine condition is fulfilled by this objective within one part in 3500, so that the coma correction is practically perfect. The analytically determined astigmatism of the second solution is in the over-corrected sense just about to the right extent to yield the most favourable approximation to a flat field. In all these respects the objective is therefore an exceptionally good one, although it departs to a startling extent from the stereotyped Lister type, which would have $r_3 = r_6 = \infty$, and very short radii of the first and fourth surfaces.

The achromatism next calls for investigation. Using the $\Sigma(d-D)dN$ method, the X of the six surfaces have first to be computed; by the equation $X = (PA)^2/2r$ they are found as

$$X_1 = \quad \cdot 01861 \qquad X_4 = \quad \cdot 01188$$
$$X_2 = -\cdot 17330 \qquad X_5 = -\cdot 04814$$
$$X_3 = -\cdot 07682 \qquad X_6 = -\cdot 02029$$

$$X_2 - X_1 = -\cdot 19191 \qquad X_5 - X_4 = -\cdot 06002$$
$$X_3 - X_2 = +\cdot 09648 \qquad X_6 - X_5 = +\cdot 02785.$$

As $D = (d + X - X_0)\sec U$, and as the values of d_1 in proper order were ·4, ·2, ·3, and ·2, the calculation proceeds

	$d+X-X_0$.	U.	$D=(d+X-X_0)\sec U$.	$d-D$.
First lens	·20809	−0° 17′ 44″	·20809	+·19191
Second lens	·29648	−1 57 55	·29669	−·09669
Third lens	·23998	8 38 32	·24044	+·05956
Fourth lens	·22785	2 22 18	·22804	−·02804

The sum $\Sigma(d-D)dN$ can now be worked out for any available glasses of the assumed refractive indices. For the glasses used in the analytical solution Chance's list gives $N_f - N_c$ for the crown as ·00856, for the flint as ·01729. Using these with the above values of $d-D$ for the several lenses it is found that

$\Sigma(d-D)dN = -\cdot 000004$ cm. $= -\cdot 08$ wave-length.

As the tolerance is ·5 wave-length, this minute over-correction may be allowed to pass. If the residual had proved larger it would have been necessary to search the glass lists for a crown with slightly higher or for a flint with slightly lower dispersion. The great advantage of the $(d-D)$ method is that it is nearly always easy to establish sufficiently exact achromatism by merely ringing the changes on available meltings, provided of course that the rough design was arrived at with due regard to approximate fulfilment of the chromatic condition.

The test for zonal chromatic aberration by the paraxial form of the $(d-D)$ equation is next carried out by the formula given above. It was found to give the paraxial residual $= +\cdot 000048$ cm., or under-correction amounting very nearly to one wave-length. As the "tolerance" was fixed at two wave-lengths the objective is found to be satisfactory.

Zonal spherical aberration must then be determined by the trigonometrical OPD_m equation (10**). Employing 20,000, the approximate number of wave-lengths in one centimetre, as an additional factor so as to obtain the OPD directly in approximate wave-lengths, the results are

$$OPD_1 = -\ \ \cdot 35 \qquad OPD_4 = +\ \cdot 005$$
$$OPD_2 = +14\cdot 84 \qquad OPD_5 = +3\cdot 32$$
$$OPD_3 = -11\cdot 04 \qquad OPD_6 = -6\cdot 12,$$

and their algebraical sum $= +\cdot 66$ wave-length represents the lead with which the

marginal ray meets the axial ray at the final focus. As the present system includes no very large angles or curvatures this residual may safely be interpreted as indicating simple zonal aberration, for which a tolerance of two wave-lengths was established. The result is therefore highly satisfactory.

It may be of interest to add that the whole of the analytical and trigonometrical work involved in arriving at this final design was carried out in less than ten hours The objective would be described as a $2\frac{1}{2}$ inch of NA. 135, for English tube-length. If made to two-third scale it would be suitable for a $1\frac{1}{2}$ inch for Continental tube-length.

In the higher powers of three or more separated components the attainment of the desired correction depends almost entirely on empirical trials guided by experience and instinct, the latter being most easily acquired by progressive studies beginning with simple achromatic lenses and gradually advancing to more complex forms. The calculation of the OPD values should never be omitted, for experience has shown that if the differences of optical paths arising at any one surface exceed 30 or at most 50 wave-lengths the fulfilment of the Rayleigh condition for the whole aperture is rendered almost impossible. Similarly the sum of all the positive OPD amounts should not exceed 100 or at most 200 wave-lengths for any one complete system.

In systems with the usual thick plano-convex front lens the form of the latter which leads to correction of the longitudinal spherical aberration and of the coma for the computed marginal ray can always be determined by a practically direct solution. Refraction at the final plane surface does not change the ratio of sin U′ to u', as both are derived from their values before refraction by multiplication with the relative refractive index. Therefore the sine condition must be fulfilled by the U′ and u' of the first, spherical face of the front lens, and the requisite radius can be quickly found by a few trials based on this criterion. Provided that this radius leads to spherical over-correction of the refracted pencil (if it does not, then spherical correction will be impossible) the necessary thickness of the front lens can be found by a direct solution. Let the intersection lengths of the rays issuing from the convex surface be l' and $L' \equiv l' + (L' - l')$, and let d be the sought thickness. The intersection lengths of the rays arriving at the plane face will be $l' - d$ and $l' - d + (L' - l')$ respectively. The corresponding intersection lengths after refraction are to be equal (so as to remove the spherical aberration) and are determined by $Pl(3p)^*$ and $Pl(3)^*$ respectively. Hence the condition

$$\frac{(l'-d)N'}{N} = [(l'-d)+(L'-l')]\frac{N'\cos U'}{N\cos U},$$

in which U represents the obliquity of the marginal ray arriving at the plane and U′ its value after refraction as obtained from the equation by $\sin U' = \sin U . N/N'$. This equation immediately gives the solution

$$d = l' - \frac{(L'-l')}{(\cos U/\cos U') - 1}.$$

With this convenient solution any number of front lenses can be found for given back combinations and the best one picked out by the criterion of zonal aberrations. This represents the strongest argument in favour of calculation of microscope objectives in the reverse direction.

In objectives of a numerical aperture exceeding about ·25 the spherical aberration of the cover-glass usually employed in microscopy has to be allowed for. As objectives of these higher powers always have a plano-convex front lens this allowance is easily made without any laborious calculation by diminishing the thickness of the front lens calculated by the last equation by the thickness of the cover-glass. This simple procedure is legitimate even when the refractive index of the front lens differs very considerably from that of the cover-glass, for it is found by direct calculation that the aberration of a thin plano-parallel plate is almost independent of its refractive index in the region of usual glass indices.

§ (21) Eyepieces.[1]—The image produced by a microscope objective is usually further magnified by an eyepiece inserted in the upper end of the tube.

The cones of rays issuing from the objective towards the primary image have a common base at or near the back lens of the objective; it is convenient and usual to assume that this common base of all the image-forming pencils coincides with the second focal plane of the objective. Any one of these pencils is refracted by the lenses of the eyepiece as indicated in *Fig.* 16, and should, for a normal eye with

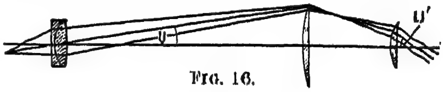


Fig. 16.

relaxed accommodation, emerge as an inclined bundle of parallel rays. The angular magnification of the eyepiece is then defined as the ratio tan U′/tan U, U being the angle at which the central or principal ray of the oblique pencil leaves the second focal point of the objective, and U′ the corresponding angle after passage through the eyepiece. The angular magnification so defined represents the number engraved on many eyepieces—more especially on compensating ones—and is the most convenient one for the designer. From the general theory of lenses it follows that if D represents the

[1] See also "Eyepieces."

distance from the upper focal plane of the objective to the lower focal plane of the eyepiece (=the so-called optical tube-length of the microscope), f the equivalent focal length of the eyepiece, and MA the angular magnification as defined above, then

$$\text{MA}=\frac{-\text{D}}{f}.$$

The values of D used by Abbe for the compensating eyepieces have been widely adopted; they are D=180 mm. for the "Continental" and D=270 mm. for the "English" standard tube-length. English makers, however, more usually adopt 10 inches as the value of D on which their magnification numbers are based.

The order of importance of the various aberrations is quite different for eyepieces from that for objectives. The reason is that the individual image-forming pencils passing through the eyepiece are very slender, their convergence ratio as they leave the objective rarely reaching even $f/16$. Consequently the longitudinal spherical and chromatic aberration of the individual pencils is unimportant and may usually be ignored altogether. On the other hand pencils like the one shown in *Fig.* 16 aiming at image-points near the margin of the field of view pass the eyepiece lenses at large angles of incidence (frequently exceeding 30°, and at the contact surface of achromatised eye-lenses even 50°), and these large angles may lead to very serious amounts of chromatic difference of magnification, of distortion, of astigmatism, and even of coma. These are the aberrations which call for the attention of the designer.

The chromatic difference of magnification is most easily dealt with. It arises from the dispersion of a white principal ray into its constituent colours by the prismatic effect of the extra-axial parts of the eyepiece lenses. Such a ray will be dispersed by the first or field-lens (a) as indicated in *Fig.* 17, and if a

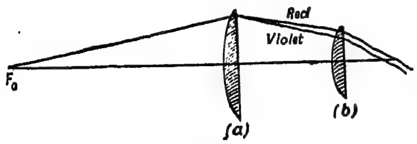


FIG. 17.

single lens were used as an eyepiece the violet and red images would be seen in the reverse direction of the corresponding emerging rays and would not coincide: the red image would appear decidedly smaller than the violet one. This defect is most usually corrected by placing a second lens, the eye-lens (b), at a considerable distance from (a). The violet ray then traverses this second lens closer to its centre than the red ray and suffers less deviation, and a means is provided to secure parallel emergence of the different coloured rays. As was pointed out in discussing the objective, the latter may itself supply primary images with a chromatic difference of size, the violet primary image being larger than the red one. The eyepiece then must compensate this difference in addition to correcting its own. The following paraxial solution is usually sufficient to determine the data of an eyepiece consisting of two thin lenses so as to establish equal magnification in all colours:

Let C be the compensation constant of the objective with which the eyepiece is to be used, calculated as C=1 − (size of red primary image)/(size of blue primary image), the red image corresponding to the C-line of the spectrum, the blue image to the F-line. For most of the apochromatic objectives C=·012. In the case of eyepieces for general use with ordinary objectives it would be advisable to introduce C=·004 or ·005 as a compromise.

Let f be the desired equivalent focal length of the eyepiece, f_a the focal length of the field-lens, f_b that of the eye-lens, $k=f_b/f_a$ the adopted ratio of the two separate focal lengths, a usually suitable numerical value being $k=\cdot 5$.

Further let ν_a be the glass-maker's ν-value for the material of the field-lens, ν_b the corresponding value of the eye-lens, MA the required angular magnification, always a negative number, and d the separation (air-space) between the two thin lenses. Then the problem will be solved by calculating

$$f_b=f[\text{P}\pm\sqrt{\text{P}^2-\text{Q}}],$$

in which

$$\text{P}=1+\tfrac{1}{2}\text{MA}\,.\,k\left(\frac{\nu_a+\nu_b}{\nu_b}+\text{C}\,.\,\nu_a\right),$$

$$\text{Q}=(1+\text{MA}\,.\,k)\left(1+k\frac{\nu_a}{\nu_b}\right),$$

and then

$$f_a=\frac{f_b}{k},\quad\text{and } d=f_a+f_b-\frac{f_a f_b}{f}.$$

For eyepieces without compensation (C=0) two plano-convex lenses of ordinary crown-glass will usually give fairly satisfactory results. If the Huygenian type is to be adhered to for compensating eyepieces, good results are only obtainable by making the field-lens of dense flint-glass (Nd=1·62 or better 1·65) and the eye-lens of light crown-glass, or better still by using an achromatised eye-lens ($\nu_b=\infty$). For compensating eyepieces for high values of NA Abbe introduced chromatically over-corrected close combinations on account of the greater distance of the eye-point obtainable with this type. With Huygenian eyepieces of high magnification the eye has to be brought very close to the eye-lens

in order to see the whole field, and this proves irritating, especially to observers with long and stiff eyelashes.

The modification of an eyepiece design with a view to reducing the other aberrations mentioned involves the whole theory of pencils of finite obliquity and aperture and cannot be usefully dealt with in the available space.

§ (22) THE ILLUMINATION OF MICROSCOPIC OBJECTS.—From the point of view of geometrical optics it ought to make very little difference how the illumination of an object is effected, provided it is of suitable intensity for comfortable observation. If aberrations are absent the instrument should yield a perfectly sharp image of every point in the object regardless of whether the whole aperture of the object-glass is filled with light or whether only a part of the aperture transmits light. Moreover, such an absolutely sharp image should bear unlimited magnification. This was in fact the view taken until about 100 years ago; if the instrument failed to give sharp images when a certain magnification was exceeded, the explanation was sought in uncorrected aberrations. Certain observational experiences which demonstrated that with the same instrument and the same magnification sharper images and higher resolving power could be obtained sometimes merely by changing the direction from which the object was illuminated were explained as shadow-effects of the type which renders visible under oblique illumination countless small craters on the moon of which not a trace can be seen in the full moon. The improbability of any appreciable shadow being produced by the usually frail and highly transparent microscopic objects was ignored.

The undulatory theory of light supplied the means of explaining this obvious breakdown of the purely geometrical explanation of optical images. Airy's determination of the form and size of the spurious disc produced by interference of the light arriving at the geometrical focus of a perfect instrument solved the problem to a considerable extent. Although his result was obtained nominally for the telescope it is equally applicable to any other perfectly corrected optical instrument, for the diameter of the spurious disc is found to depend only on the wave-length of the light (to which the diameter of the disc is proportional) and on the aperture of the object-glass, if the diameter of the disc is measured in angle, or on the ratio of aperture to focal length if the diameter of the disc is measured in linear measure; the diameter of the disc in either case is inversely proportional to the aperture. As two discs will cease to be separated by a dark space when their distance apart is equal to their diameters it follows that the least distance at which two points in the image can be resolved is equal to the diameter of the spurious disc, no matter in what unit the latter is measured, and also that this distance is inversely proportional to the effective aperture of the object-glass and directly proportional to the wave-length of the light employed.

When applied to the microscope, Airy's theory of the spurious disc explains some of the observed phenomena fairly satisfactorily. As the smallness of the image depends on the effective aperture, *i.e.* that part of the lenses which is really filled with light, there should be a maximum of resolving power when the object-glass is completely filled with light and a reduced resolving power if only a part of the full aperture is utilised. This is qualitatively in accordance with experience, but not quantitatively, for it is found that a microscope objective retains one-half of its maximum resolving power if only a very small axial illuminating pencil is employed; by the theory the resolving power should be reduced to a very small fraction of its maximum value. Light scattered by the structure of the object and thus slightly utilising the space not filled with direct light was naturally adduced in explanation; but it is impossible thus to account for the fact that—with all kinds of objects—the resolving power is just half that obtainable from the lens when completely and uniformly filled with light.

Another and even graver objection to the application of the Airy theory in the case of the microscope arises from the fact that the theory is only valid, firstly, if all the light arriving at any one image-point is coherent, that is, derived from one original point-source and, moreover, if this light approaches the image-point as a truly spherical wave train; secondly, if other image-points in the immediate neighbourhood of the one first considered receive no light from the source illuminating the first, otherwise there would be secondary interferences between adjacent image-points and the form of the individual images would be profoundly modified. These fundamental conditions could only be fulfilled if the objects under observation were rendered self-luminous, say by making them white hot, which is manifestly out of the question in all but most exceptional cases. It used to be thought (Helmholtz was the last responsible scientist upholding this view) that the objection could be met by projecting upon the object under examination the sharp image of a self luminous source by means of a highly corrected condenser. But no condenser yet made has the requisite freedom from aberrations, and even if one were produced it would still depict each point of the source as a tiny spurious disc with attendant interference-rings, all consisting of coherent light covering an appreciable area

of the object and so defeating the theoretical requirements.

Finally the Airy theory can give no reasonable explanation of any kind of the results obtained by lighting objects with light of such obliquity that the direct illumination cannot enter the objective of the microscope, that is, the theory fails entirely for dark-ground illumination.

§ (23) ABBE THEORY.—Abbe was the first physicist who clearly realised these shortcomings of the Airy theory when applied to the microscope. He overcame the difficulty in the way which is rendered obvious the moment its nature is recognised, by studying the nature of the image produced when the object is illuminated by a single point-source so that all the light falling upon the object is necessarily coherent and capable of interference. If an extended self-luminous source is subsequently substituted, each point of it will produce an image of the type determined, and as the vibrations of different points of a self-luminous source are independent and incapable of permanent interference the final image will everywhere have the simple sum of the intensities of the elementary images without any complications due to secondary interferences.

If a minute hole measuring only a small fraction of a wave-length in diameter is produced in an opaque film and illuminated from a distant point-source, then there will be no scope for sensible interference effects, and the hole will send out light in all directions on the principle of Huygens and will behave exactly as if it were itself self-luminous. This is the only case in which the theories of Airy and of Abbe lead to the same result. It is approximately realised in the "ultra-microscope," by which minute particles are seen exactly like stars in a telescope; in fact, if the instrument is well corrected the spurious disc with its surrounding diffraction rings is clearly seen.

But if the light-transmitting aperture attains a size of the order of a wave-length or if there are a number of apertures within small distances of each other, then there will be interference effects between the light from different points of the aperture or apertures, resulting in the producing of some type of diffraction spectra. In the vast majority of cases the result is hopelessly complicated and, moreover, inaccessible to theoretical discussion, inasmuch as the latter would require the minute structure of the object to be known with absolute certainty. In the case of delicately structured natural objects only known by microscopical observation an obvious vicious circle would be involved in using such objects to test a theory. This also was clearly realised by Abbe and avoided by devising a large number of highly ingenious experiments on artificial simple objects such as finely-drawn-out glass threads, and especially on gratings ruled in a silver or carbon film and used either singly or two crossed so as to produce dot-patterns. By observing such objects of accurately known structure Abbe proved experimentally his theoretical conclusions that refraction or reflection of light by the object could not explain the image, but that the image could be fully accounted for by the diffraction spectra produced by the object, and that the verisimilitude of the image depended on the extent to which the diffracted light was admitted by the objective of the microscope. For a fuller account of this Abbe theory reference must be made to Dippel's handbook of microscopy, and for a more detailed study of the application of the theory to the conditions usually prevailing in the actual use of the microscope to two papers by Conrady.[1]

All that is required to establish the theory of illumination is to consider the distribution of the diffraction spectra in the case of a simple grating of straight and equidistant slits. If such a grating is illuminated by parallel light from a distant point-source a number of diffracted beams are produced, and as the angles at which these proceed obey a sine law whilst a microscope objective also must satisfy the optical sine condition, it easily follows that, no matter in what direction the source of light may be placed, the diffracted beams are depicted in the upper focal plane of the objective focussed on the grating as a series of equidistant points if the light is monochromatic, the distance from point to point being inversely proportional to the spacing of the grating and to the wave-length of the light, and the line formed by all the diffracted light-points being at right angles to the direction of the grating-slits. If at least two consecutive diffracted beams are transmitted by the objective, then their interference in the final image plane (usually on the retina of the observer's eye) will produce an alternation of bright and dark lines coinciding with the image of the grating which would be deduced by geometrical optics. With an objective of limited aperture only a part of the complete set of diffraction spectra will be admitted. Referring to *Fig.* 18, in which the direct light is distinguished by a small circle, the flanking diffraction spectra by crosses, it is easily seen that if the source of light were placed upon the optical axis, an objective of small aperture (A) might admit only the direct light and no resolution would be obtained. It would require an aperture B to admit the direct light and the two nearest diffraction

[1] *Journ. R. Micr. Soc*, 1904, pp. 610-633, and 1905, pp. 541-553.

spectra with axial illumination and to secure resolution. But if the source of light were moved to one side of the optical axis—thus

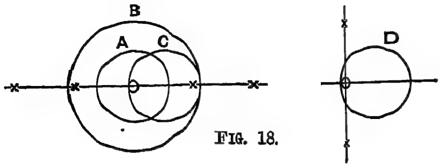


FIG. 18.

producing oblique illumination—then the aperture A would just be capable of admitting the direct light and the nearest diffraction spectrum (C) and resolution of the grating would result. The increased resolving power with oblique illumination is thus explained; also the two-to-one ratio of resolving power with extremely oblique and with axial illumination, respectively.

But a very grave objection to simple oblique illumination is also easily deduced. As a rule the observer will be ignorant of the direction of the lines of the structure under observation or there may be in the same field specimens having lines in various directions. Now with oblique light arranged according to case C and a structure producing a row of diffraction spectra of the same, or even a closer, spacing, but on a line nearly at right angles to that first considered, only the direct light would be admitted (D) and no resolution would result. To avoid this pitfall it is therefore absolutely necessary when employing simple oblique illumination to try it in all azimuths, either by rotating the object by means of a mechanical stage or by producing a corresponding change in the illuminating pencil by a rotating eccentric diaphragm of the substage-condenser. But a method which nearly always deserves preference is to produce oblique light in all azimuths simultaneously by means of a condenser opened to a suitable aperture to secure the desired degree of obliquity with a large "solid cone" of illumination, or to accentuate the higher resolving power of oblique light by adding a central stop below or above the condenser-iris which cuts out the central part of the illuminating cone and so produces "annular illumination." Evidently the essential requirement is that the illumination should be perfectly symmetrical with regard to the optical axis, that is, strictly concentric with the latter. Looking down the tube of the microscope the observer should see either a uniformly bright and nicely centred disc of light or a ring of light of uniform brightness and bounded by circles concentric with the axis of the microscope. These are the conditions which are both necessary and sufficient to secure truthful and reliable images of microscopic objects no matter what their structure may be. The realisation of these conditions is not difficult if a uniformly bright source of light of large area, such as the sky or a large flame, is available, and excellent results are then obtainable with condensers of the simplest type. But when the source is small and especially when it is almost linear like an incandescent lamp filament, then the conditions can only be adequately satisfied with a condenser carefully designed so as to be fairly free from spherical and chromatic aberration and accurately focussed so as to project a sharp image of the source of light upon the object under observation. Condensers of this kind closely resemble microscope objectives of corresponding numerical aperture and are designed by the methods described in earlier sections: the tolerances, however, may be safely extended to two or three times the amounts stipulated for objectives.

The Abbe theory also leads directly to a formulation of the proper conditions to be fulfilled in dark-ground illumination. If a regular structure is illuminated by light of such obliquity that the direct light cannot enter the microscope objective, then the image will be produced entirely by diffracted light and at least two diffraction spectra must be admitted in order to secure an image of the structure. *Fig.* 19 shows at once that the maximum of resolving power will be secured if the first and second diffraction spectra can just enter at opposite margins of the aperture. Evidently the direct light will then be three times as far from the axis as either diffraction spectrum, and this maximum can therefore only be realised if a condenser is available which has three times the numerical aperture of the objective to which it supplies dark-ground illumination. As the numerical aperture of a condenser cannot exceed 1·5, it follows that the full resolving power of objectives over ·5 NA cannot be realised by dark-ground illumination; the latter therefore implies a disadvantage in the case of all objectives of high NA. On the other hand, there may be present in the same field coarse structures which give closely spaced diffraction spectra. In order that the latter may be admitted, it is necessary that the ring of dark-ground illumination should extend inwards as nearly as possible to the aperture of the objective in use. Hence a second important rule: that the dark-ground stop should be only just large enough to secure a dark field. This rule cannot usually be satisfactorily fulfilled if a high-power condenser has to be used

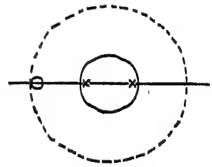


FIG. 19.

in order to secure illuminating cones of great obliquity. The reason is that in practically all existing condensers of high N.A. the diaphragm and dark-ground stop are placed far below the anterior focal plane, with the result that a real image of both is formed close above the object under observation (*Fig.* 20). If this

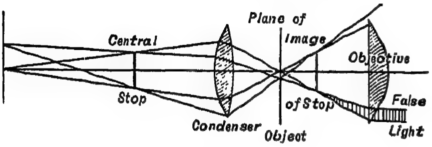


FIG. 20.

image of the dark-ground stop is only just large enough (as it ought to be) to stop direct light from the axial object-point from entering the microscope objective, then it will stop out an *oblique* cone of similar angle from extra-axial object-points, with the result that from the latter direct light will be admitted on the side of least obliquity of the inclined hollow illuminating cone. The field thus is dark only in and near its centre, with rapidly growing illumination towards its margin, and the effect is totally spoilt. This grave drawback can only be removed by placing the diaphragm and central stop close to the anterior focal plane of the condenser: unfortunately the present design of microscope stages and substages renders the fulfilment of this important condition very difficult. The various types of reflecting condensers (paraboloids, etc.) are attempts to overcome this difficulty: it would be decidedly preferable to have oil-immersion condensers with correctly placed diaphragms, and it is therefore to be hoped that designers will give earnest attention to this desideratum.

A. E. C.

MICROSCOPE EYEPIECES. See "Eyepieces," § (6); also "Microscope, Optics of the," § (21).

MICROSCOPE OBJECTIVES, CALCULATION OF. See "Microscope, Optics of the," § (2) *et seq.*

Testing of. See "Objectives, Testing of Compound."

MICROSCOPIC OBJECTS, PROJECTION OF. See "Projection Apparatus," § (16).

MICROSCOPY WITH ULTRA-VIOLET LIGHT

§ (1) INTRODUCTION.—The evolution of the modern microscope has been a process in which improvements of importance have alternated with long periods of apparent inactivity. It has in general followed the same lines as other optical instruments, that is, so far as methods of correcting for spherical and chromatic aberration are concerned (see "Microscope"). The introduction of immersion lenses by Amici in 1855 marks one of the great optical improvements that has had a profound influence on its development, although it was at a much later date that full advantage was taken of the method. Not until the latter part of the last century, largely owing to the researches of Abbe,[1] were the factors governing resolving power recognised and understood. It was realised as the result of those researches that there exists a limit beyond which resolving power cannot be increased. The practical result therefrom was the computation and production of apochromatic objectives, in which colour correction was of a higher order than had at that time been achieved in any optical instrument. Since that period there has been no real advance in microscopic optics.

There are two distinct methods of investigation in modern microscopy, one in which enhanced visibility is the main purpose in view, no information as to the form or structure of minute objects being secured, and the other the more usual one of the resolution of structural details. In the former case objects are rendered self-luminous, appearing as bright bodies on a dull ground. The method is known as "dark-ground illumination" when bodies within the resolution limits are being observed, as "ultra-microscopy" when small particles beyond these limits are made visible.[2] Considerable confusion has arisen owing to the use of the term "ultra-microscopy" in describing both modes of observation. In their fundamental physical aspects the method of dark-ground illumination and observation with the ultra-microscope are identical, as both serve to enhance the visibility of an object. The practical difference is in the type of object under observation and the manner in which it is observed, rather than in the difference between the character of the appliances used for the purpose. When objects seen in a dark field show structural details or well-defined contours we are dealing with simple dark-ground illumination. When, on the other hand, the field is seen to contain particles in which there is no trace of detail or evidence of structure, and which appear as mere points of light, the case is one of ultra-microscopic observation. It is obvious that both types of image may be seen at the same time; the fundamental difference is therefore dependent on the size

[1] Abbe, *Archiv f. mikr. Anat.*, 1873, ix. 413; *Journ. Roy. Micr. Society*, June 1881, p. 388.

[2] Siedentopf, *Journ. Roy. Micr. Soc.*, 1903, p. 573; *Ann. d. Phys.*, 1903, pp. 1-39; *Berl. klin. Wchnschr.*, 1904; *Zs. f. wiss. Mikros.*, 1907, p. 382; Faraday, *Proc. Roy. Inst.*, 1854, p. 310.

of the object, whether it is above or below the resolution limits, and on the distance between neighbouring visible units. It is in the direction of obtaining increased resolution that the greatest need is felt in practical microscopy.

§ (2) Resolving Power.—It is well known that the smallest distance separating two points in an object—such as a grating—is given by the relation

$$d=\frac{\frac{1}{2}\lambda}{\sin a},$$

where λ is the wave-length of the light used in the medium surrounding the object, and a is the angle subtended at any point in the object by half the effective angular aperture of the objective.[1] This is usually written

$$d=\frac{\frac{1}{2}\lambda_0}{\mu \sin a},$$

when λ_0 is the wave-length in air, and μ the refractive index of the medium. It follows that the value of d, the distance between any two elements of recurring structure, is dependent on the wave-length of the light used and on the product of $\mu \sin a$, which is known as the numerical aperture (N.A.) of the illuminating and observing optical system. This relation is applicable in the case of a grating illuminated by a beam of the greatest possible obliquity, ensuring that the essential diffraction spectra are within the limits of the objective aperture. It is reduced to

$$d=\frac{\lambda_0}{\mu \sin a},$$

when axial illumination is resorted to, a condition that applies to most observational work. It is in this connection important to remember the well-founded view of Abbe that "the diffracted light emanating from the object may utilise the whole aperture of the system, although the incident cone of light, if it were simply transmitted in the absence of an object, would fill only a very small portion of the aperture." It follows that greater resolution can be obtained either by reducing the wave-length of the light used as the illuminant or by increasing the N.A. of the objective. With dry lenses the greatest attainable N.A. is ·95. This can be increased by using immersion lenses, the resulting N.A. in the case of water being approximately 1·25, cedar-wood oil 1·50, or monobromide of naphthaline 1·60, the latter being used only in metallurgy, or in cases where flint cover-glasses of the necessary refractive index can be utilised. The actual resolution obtained with white light with a mean wave-length of 550 $\mu\mu$, when a grating is illuminated by oblique rays and with N.A. 1, is therefore about 92,000 lines per inch. The use of oblique light is necessary to obtain the utmost resolution with such an object, as otherwise the diffraction spectra which are essential to the formation of the image cannot be embraced by the objective. It has to be recognised that such an image is not of necessity a true representation of the object; it is in most cases no more than an indication of the periodicity of the structure. The so-called resolution of diatoms may be cited as an example, in which the structure may give rise to an image consisting of parallel lines or bars. Such lines are known to be merely an indication of a regular sequence of depressions, perforations, or excrescences of the same periodicity, and this constitutes the only point of similarity between the object and the image. To obtain a real image the object must be illuminated with a solid axial cone of light, and under these conditions resolution falls to one-half of the value stated. It would therefore be of the order of 46,000 lines per inch in white light. Johnstone Stoney[2] states that with an immersion objective of N.A. 1·35, and an immersion condenser of N.A. 1·30, using a suitable stop to secure oblique illumination in a direction at right angles to the direction of the lines of the grating and with blue light of 450 mm. wave-length, the practical limit of separation is about 0·20 μ. He has also demonstrated that a pair of objects may be resolved when they are separated by about five-sixths of the interval between the elements of a row of such objects or points; in other words, that resolution is increased appreciably when only two neighbouring elements of structure are being observed. The computation by Abbe[3] of apochromatic objectives has resulted in the production of systems for visual work with the highest attainable N.A., and in which spherical and chromatic correction are of a high order. Their chief characteristic is that they are perfectly corrected for three colours in the spectrum, and approximately for all colours in white light. Their superiority over any achromatic system is dependent on their computation and not on the use of any particular material, such as fluorite, in their construction, although this material has optical constants which are essential in the computation of such objectives, no suitable optical glass yet being available. By using light of short wave-length in the spectroscopic region of the green, blue-green, or blue-

[1] Abbe, *Abhandlungen über die Theorie des Mikroskops*, 1904; Rayleigh, Lord, *Phil. Mag.*, 1896, p. 467; *Journ. Roy. Micr. Soc.*, 1903, p. 447; Porter, *Physical Review*, 1905, p. 386; Glazebrook, *Trans. Optical Soc.*, 1907–1908, p. 94. See also "Microscope, Optics of the," Introduction.

[2] Johnstone Stoney, *Journ. Roy. Micr. Soc.*, 1903, p. 564.

[3] Abbe, *Journ. Roy. Micr. Soc.*, 1879, p. 377.

violet, the greatest attainable visual resolution is secured. The most useful illuminant for this purpose is the mercury-vapour lamp, as it is possible by using suitable colour screens to utilise the bright lines in the orange, green-blue, or violet. Even when used without a screen this illuminant has the advantage possessed by no other, that the mean wave-length is reduced, as no red radiations enter into its composition. It is obvious that even under these conditions resolution is strictly limited. The only practical method of obtaining increased resolution that has been evolved since the researches of Abbe, with proportionately increased magnification, which constitutes a definite advance, is in the use of ultra-violet light, a method which opens up a wide field for investigation. The shortest wave-length that has sufficient luminosity in the visual spectrum for direct observation may be regarded as being in the region of wave-length 475 $\mu\mu$, but if it is possible to utilise ultra-violet radiations of a wave-length as short as 220 $\mu\mu$, resolution will, other things being equal, thereby be doubled. There is the self-evident objection to such a method, that the image can no longer be a visual one, and, further, that the optical system must be composed of materials which are transparent to ultra-violet light.

§ (3) Quartz Optical Systems. — The transparency of any optical glass is not sufficient for work with the more refrangible portion of the spectrum, even a borosilicate crown-glass transmits but a small percentage of light of wave-length 300 $\mu\mu$. Schott of Jena has made optical glass particularly transparent to the ultra-violet light, but at wave-length 280 $\mu\mu$ a specimen 1 mm. thick only transmits 50 per cent of the incident light, and with shorter wave-lengths it is practically opaque. The only materials available therefore are quartz, either crystalline or fused, and fluorite. Fused quartz is rather less transparent than crystalline, but in carefully selected pieces which are free from impurities the difference for the purpose in view is not important. Fluorite was first used in the construction of microscope objectives by Spencer in America as long ago as 1860, while Boys in England, in or about 1885, suggested the suitability of fused quartz for this purpose. It was not until 1900 that M. von Rohr [1] of Jena succeeded in constructing a quartz-fluorite objective of about 4 mm., equivalent focus N.A. 0·30. This may be regarded as experimental, as it was supplanted in 1904 by a series made entirely of fused quartz. These objectives were termed "monochromats," as they are computed for a wave-length of 275 $\mu\mu$, although they can be used satisfactorily over a considerable range in the ultra-violet region, provided that the light is monochromatic. The equivalent focal lengths are as follows:

6 mm. dry lens N.A. in white light 0·35.
2·5 mm. glycerine immersion N.A. white light 0·85.
1·7 mm. glycerine immersion N.A. white light 1·25.

The effective N.A. of these lenses when used with ultra-violet light may be regarded as double the values indicated, being respectively 0·70, 1·70, and 2·50, this being the approximate relation of the mean wave-length of white light to that of ultra-violet (275 $\mu\mu$). The immersion fluid is glycerine and water with a refractive index 1·447/D. The immersion fluid must be sufficiently transparent to ultra-violet light, and in this respect glycerine fulfils the purpose, but precautions must be taken in practice to ensure that its refractive index remains constant, a short exposure to atmospheric influence causing appreciable alteration.

Five oculars for projecting the image on to the photographic plate have been computed, and these are made from crystalline quartz. Their initial magnifications are 5, 7, 10, 14, and 20 when used at a mechanical tube-length of 160 mm. The maximum magnifications obtainable are set out in the following table:

Objective.	Eyepieces . . .	5	7	10	14	20
6 mm. N.A. 0·70	Optical camera length . Magnification . .	30 cm. 250	34 cm. 400	30 cm. 500	34 cm. 800	30 cm. 1000
2·5 mm. N.A. 1·70	Optical camera length . Magnification . .	30 cm. 600	30 cm. 800	31·5 cm. 1200	30 cm. 1600	31·5 cm. 2400
1·7 mm. N.A. 2·50	Optical camera length . Magnification . .	31 cm. 900	32 cm. 1300	31 cm. 1800	31 cm. 2500	31 cm. 3600

The term "optical camera length" is the distance of the sensitive plate from the upper focal point of the microscope, this being for all practical purposes the cap of the eyepiece. The substage condenser is of quartz and is made with two interchangeable top lenses,

[1] Kohler and von Rohr, *Zeitschrift f. Instrumen.*, 1904.

so that the illuminating beam may be of such N.A. as is desirable with each objective.

FIG. 1.—Microscope and Camera ready for Photographing.

A, Fluorescent searcher eyepiece. B, Fluorescent uranium glass disc. C, Right-angled quartz prism.

Without either top lens the condenser is a dry one suitable for use with the 6 mm. objective. The complete condenser is used with the immersion lenses, and is itself also immersed with glycerine. An iris diaphragm is provided as in all substage illuminators. The illuminating beam is deflected in the direction of the optic axis of the microscope by means of a right-angled quartz prism which replaces the mirror on an ordinary microscope. The essential parts on the microscope are therefore the same as for visual work, except that they are all of quartz.

Figs. 1, 1A, and 2 show the microscope and camera and the arrangements for illuminating the object respectively.

§ (4) OBJECTS.—The object to be photographed must be mounted on a quartz slide and covered with a quartz cover-glass. The object-slides are of crystalline quartz, ground at right angles to the optical axis, and are 0·5 mm. thick.

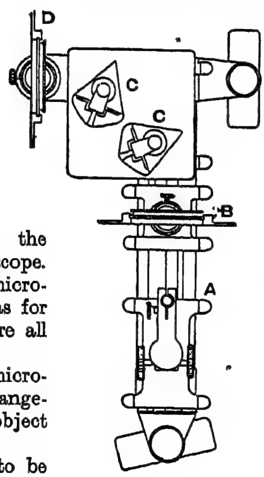


FIG. 2.

The use of optically worked slides and cover-glasses of definite and constant thickness is essential, as otherwise the difficulty of adjusting the apparatus would be greatly increased. In visual work the adjustment of tube-length to correct for spherical aberration introduced by cover-glasses of varying thickness is essential, but this would be a matter of considerable difficulty

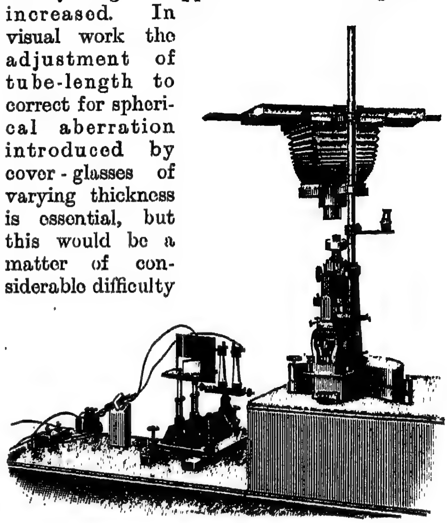


FIG. 1A.—Showing the whole Apparatus in Action.

when using ultra-violet light. For the same reason it is essential that the object should be immersed in as thin a layer of mounting medium as possible, otherwise the effect on the performance of the objective is the same as that resulting from the use of a thicker cover-glass.

The object, or element of structure under examination, must in fact be nearly, if not quite, in contact with the under surface of the cover-glass. There are certain limitations as to the type of the object that can be dealt with by this method—limitations which are determined in large part by the transparency of the object to ultra-violet light. The term "transparency" used in this sense is a relative one, as many substances that in thicker layers are completely opaque become very translucent when in thin microscopic layers. Some micro-organisms, for instance, which in the direction of their breadth are beyond the limits of microscopic resolution, have entirely baffled all efforts to photograph them by means of ultra-violet light. The radiations pass

completely through, and no image is formed on the photographic plate. This, however, is probably a question of wave-length. Radiations of shorter wave-length, when they can be utilised for the purpose, will probably do all that is required, and these very small organisms can then be dealt with. The objects best suited for this method are bacteria, yeasts, or any cellular structure which can be observed as separate units or in thin layers. Apart from the difference obtained in the image as the result of differences of thickness, there is a differentiation of structure due to differences of organic constitution which is shown in the resulting photograph. Spore-forming organisms, for instance, result in a somewhat striking picture; the organism itself may be somewhat translucent, but the spores are almost invariably quite opaque to ultra-violet light, although they are of similar dimensions. The same applies to granular contents in yeast cells and other cellular structures of a similar character.

The method therefore acts as a staining reaction without the necessity of carrying out any processes which might result in alteration in the appearance of the structural elements. Objects must in all cases be unstained. They must not be treated by any of the well-known methods used in bacteriology or cytology, in which fixation, hardening, or drying takes place. These processes probably cause some change in organic tissues, and in most cases they result, in the method under discussion, in the tissues becoming opaque to the radiations used. It is obvious therefore that the results are a representation of the tissues or micro-organism under actual living conditions.

In practice the preparation of the object presents no difficulty. The only point is that it must be mounted between its cover-glass and slide in some fluid which is transparent to ultra-violet light. Such fluids are not numerous; but distilled water, glycerine, or glycerine and water, Ringer's solution, normal saline, or castor-oil are among the substances available. To inhibit motion in living organisms a ½ per cent solution of agar may be used, or a dilute solution of gelatine. This procedure is, however, to be avoided if possible, and it is preferable to rely on thin layers of material so that motion is inhibited by the short distance, and consequent small free path, between slide and cover-glass. The ordinary nutritive media, as used by bacteriologists, are in some cases suitable, but they may be diluted to about one-fifth or one-sixth their ordinary strength. Sections may be passed from xylol to vaseline oil or castor-oil, but it is necessary to see that the xylol is completely removed, as that substance is opaque. In general, scrupulous cleanliness is essential, as otherwise the dust particles which occur in nearly all preparations become all too evident. The quartz slides and cover-glasses must be boiled in bichromate of potash and sulphuric acid to ensure perfect cleanliness, and should then be heated in a Bunsen flame before they are put into use.

§ (5) Light Sources.—The source of light is a high-tension spark discharge between metallic electrodes, usually magnesium or cadmium. The necessary characteristics of the light are, that it should emit ultra-violet radiations in sufficient quantity, and that its spectrum should mainly consist of a few intense bright lines, or groups of lines. The lines must not be too close together, so that any individual line may be isolated by spectroscopic methods, and utilised as the source of light.

Few metals conform to the necessary conditions. Iron, for instance, is quite unsuitable, although it emits ultra-violet rays freely; but the large number of lines in its spectrum, and their closeness together, make it impossible to utilise it. Cadmium and magnesium are the most suitable, their spectra consisting of a few intense well-separated lines. Recent experiments indicate that silver and beryllium may prove of value where intervening wave-lengths are required. The length of spark required is about 5 mm., but it must be what is usually referred to as a "fat" type of spark of high intrinsic brilliancy. It appears that there is a definite limit to the amount of energy that can be used in this direction. The conditions are comparable to that of an arc, in which increase of current results in a greater volume of light owing to the increased size of the luminous crater, but does not materially alter the intrinsic brilliancy. The coil used is therefore one that gives a heavy discharge at a relatively low voltage. An oil-immersed condenser, which has been found superior to a Leyden jar, is connected in parallel with the spark, with the result that a rapidly alternating discharge takes place between the metal points. Either an electrolytic or a mercury break may be used as an interrupter; but it has been found in practice that the latter gives better results, as it runs for a long period with greater regularity.

§ (6) Methods of Observation.—The spectroscopic arrangements for projecting the necessary monochromatic light into the microscope follow the lines in general of an ordinary spectroscope, with the exception that there is no slit. In front of the spark, and at a suitable distance from it, a quartz lens of about 18 cm. focal length is placed, so that an image of the spark is projected at the

plane of the iris diaphragm in the microscope substage. A pair of crystalline quartz prisms of opposite rotation are arranged so that the light is decomposed, and are set at minimum deviation for the particular wave-length it is desired to use. The light passes from the prisms to the right-angled quartz reflecting prism underneath the microscope, and is, by means of that prism, deflected in the direction of the optic axis of the instrument. In the plane of the iris diaphragm on the microscope a disc of uranium glass, on which is engraved a small circle, is placed. A fluorescent image of the spark can be observed within this circle, and must be sharply focussed and accurately centred so that the ultra-violet rays pass into the substage condenser when the uranium glass is removed. When this is secured, the first condition essential for satisfactory adjustment is satisfied.

The accurate focussing of the image is the most difficult part of the process, and until considerable experience has been gained cannot be performed with certainty. There are two methods available: in one, reliance is placed on the fluorescent screen already described, but it must be admitted that this is uncertain. The difficulty is increased when lines in the spectrum are being used which are not of great brilliancy; and, in any case, the appearance of the image varies with each object.

The object must be examined beforehand by visible light, so that its appearance under these circumstances is known; then small elements of structure that are fairly easily seen must be selected for focussing purposes. If these are in the same plane as unknown structural elements the result is satisfactory, but the element of chance is too great for the process to be regarded as a reliable one.

In the second method the image is observed by means of a quartz objective and ocular, the illuminant being one of the lines, either in the region of the blue or the violet, emitted by a mercury vapour lamp. Thus by using suitable screens, which are commercially obtainable, any desired wave-length is transmitted and the image at once observed and accurately focussed on any desired plane. The illumination is then changed to the particular ultra-violet line that it is proposed to use for photographic purposes. Either by calculation or by trial and error, the difference between the focal planes of the wave-length which is used for observation and that which is to be used for photographing is determined. The fine adjustment of the microscope is then moved by this predetermined amount, and the photograph at once taken. In practice it is found that this method is quite satisfactory and yields almost invariably good results. Its accuracy is mainly dependent on the mechanical perfection of the fine adjustment on the microscope, but the ability of the observer to focus an object illuminated by light of short wave-length is also a factor. It is found advisable to use a dark slide of repeating type, so that three photographs can be taken successively on one plate without any readjustment of the apparatus. A slight adjustment of the microscope fine adjustment above and below the required focus is then made for each photograph, so that, if there is any slight lack of adjustment, either one or the other of the photographs will be satisfactory. The photographic plates used must be of fine grain and with the smallest possible quantity of gelatine on its surface. Experiments have been made with plates of the Schumann type, that is, as used for photographing the far ultra-violet; but they have not proved so satisfactory as a slow photographic plate with a minimum of gelatine and the maximum quantity of silver that the gelatine will hold. Development is carried out to obtain the greatest amount of contrast, and a developer used which will give as fine a grain as possible. No printing process does full justice to the negatives, as the images are of necessity somewhat thin and the elements of structure often result in but small differences of density. To fully appreciate the advantage of the method the original negatives must be seen.

It has to be admitted that the whole process is one of considerable difficulty; it involves considerable acquaintance with spectroscopic methods as well as those incidental to ordinary microscopy. At the same time the results that have been secured are an advance on anything that can be done by any other method. In the present state of knowledge there is no indication of any optical improvement being effected which will be a substantial advance in microscopic method, but if radiations of sufficiently short wave-length can be utilised, it is difficult to see what limits can be placed on the microscope. The illustrations chosen are of simple, well-known objects, so that the differences in appearance can be appreciated. *Figs.* 3, 5, and 7 are photographs of living organisms illuminated under the best possible conditions by means of a dark-ground illuminator. *Figs.* 4, 6, and 8 are the same organisms photographed in ultra-violet light. In each case there is a suggestion of structure in the dark-ground image, but there is no comparison when careful examination is made between the two images. On theoretical as well as practical grounds the dark-ground method is useful enough for the observation of somewhat gross structures, but it utterly fails when resolution of a high order is needed.

FIG. 3.—*Saccharomyces Pastorianus*. × 2000.

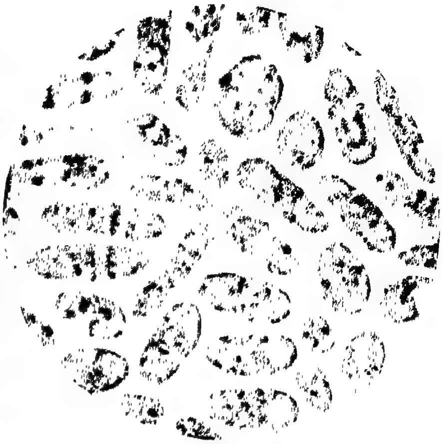


FIG. 4.—*S. Pastorianus*. × 2000.

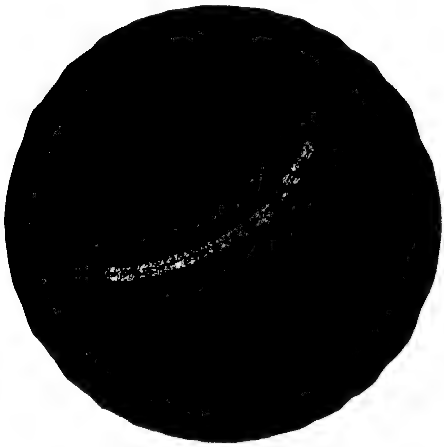


FIG. 5.—*Bacillus anthracis*. × 2000.

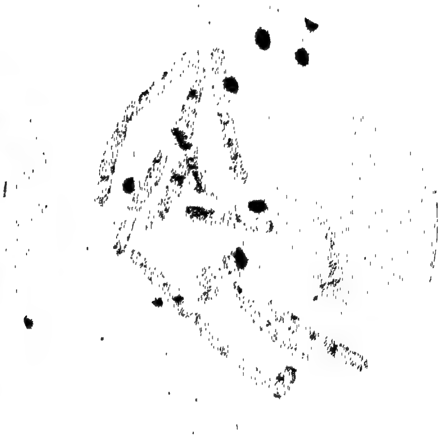


FIG. 6.—*B. anthracis*. × 2000.

FIG. 7.—*Bacillus megatherium*. × 1750.

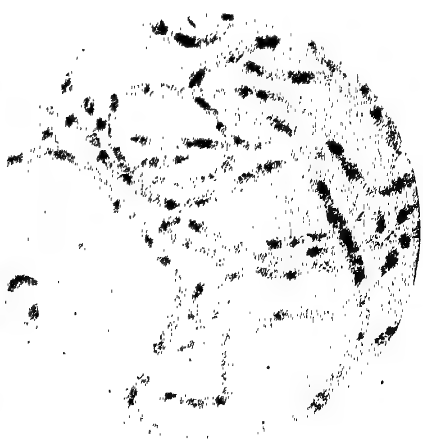


FIG. 8.—*B. megatherium*. × 1750.

§ (7) Visual Methods.—As a means of microscope research, the method so far described suffers from the disadvantage that it is purely a photographic process. The visual image obtained upon the fluorescent observing screen, which is used for the location of the image and approximate focussing, affords no indication of the ultimate value of the results. There is only one method by which a visual image can be obtained, which is indicative of structure, and that is by observing the fluorescence that occurs in certain animal tissues and other substances when they are illuminated by means of ultra-violet light. It is well known that fluorescence is brought about principally by the ultra-violet rays; the rays of the visible spectrum being much less active. The microscopic observation of the fluorescence of animal and vegetable tissues differs in practice from that hitherto described. There are two sources of light available, either an electric arc or a quartz mercury vapour lamp. The electric arc is preferable, and can be used either with ordinary carbons or with one or both carbons provided with a core of copper, iron, or nickel. Carbons impregnated with carbonate of iron also form a satisfactory light source, as the resulting radiations consist in large part of light in the ultra-violet region. The necessity for a spectroscopic arrangement to decompose the light does not arise, as the particular wave-length of the radiations employed is not of moment. In addition, the loss of light resulting from the use of a pair of quartz prisms to isolate the ultra-violet region of the spectrum is so great as to render the observation of some fluorescent images difficult, if not impossible. The fluorescence that occurs is characteristic of the fluorescing body, and is practically independent of the wave-length of the exciting light. It follows therefore that, if ultra-violet only can be used as an illuminant and all visible light excluded, a satisfactory light source is at once obtained. This isolation of the ultra-violet portion of the spectrum was not practicable until R. W. Wood of Baltimore [1] found that nitrosodimethylaniline in aqueous solution of one in five thousand to ten thousand would transmit a considerable portion of the ultra-violet spectrum while cutting out part of the visible spectrum. A quartz or uviol glass cell can therefore be filled with this solution of a thickness of about 1 cm., and this acts as an absorbing screen.[2] In addition, a piece of blue uviol glass, together with a cell containing a 20 per cent solution of copper sulphate, is used. The purpose of the sulphate of copper is to cut out the extreme red, which is otherwise transmitted, and it also absorbs invisible heat waves. A quartz lens of suitable focal length is placed in front of the arc, so that a parallel beam is projected on to a right-angled prism of quartz which takes the place of the ordinary mirror on the microscope. A quartz condenser similar to that used in the case of ultra-violet light photography is placed in the substage. The object itself, which must be in thin section, is mounted on a quartz slide in any suitable fluid which is transparent to ultra-violet light; but it may be covered with an ordinary cover-glass, as the fluorescent light emitted by the object is in the region of the visible spectrum. The only precaution necessary is to ascertain that the particular glass used does not itself fluoresce. Recently Messrs. Chance Brothers have introduced a new kind of glass which is transparent to ultra-violet light while transmitting but little visible light. If used in conjunction with a sulphate of copper screen, as already described, it forms a very efficient means of obtaining ultra-violet light, the region transmitted being between 300 and 400 micromillimetres wave-lengths. The effects obtained when unstained microscopic objects are illuminated in this way are very beautiful, and in many cases a range of colour is produced which extends over a considerable portion of the visible spectrum. In view of the subtle differences of tint that occur in the fluorescent image, the value of the result is largely dependent upon the power of the observer to detect colour differences; the colour sensitiveness of the eye is in fact the determining factor. As all solid bodies which fluoresce in ultra-violet light also have a definite phosphorescent period, it is probable that an application of Becquerel's phosphoroscope might be found to be of considerable value in these observations. Tissues are best observed in as fresh a condition as possible, and sections are preferably cut frozen to avoid the necessity of employing embedding processes, which may cause structural changes. The section itself must be thin, because if there are superimposed elements of structure the underlying ones which fluoresce cause a considerable amount of light diffusion. All animal cells and tissues, except those which are pigmented, have the power of fluorescing in ultra-violet light. Haemoglobin, or its derivatives, does not fluoresce. The following are a few typical examples of the colours which certain animal tissues exhibit as the result of fluorescence:

Unpigmented hair: intense light yellow.

Pigmented hair, such as black: no fluorescence.

Skin: intense light blue, the blood-vessels appearing dark and non-fluorescent.

Teeth: intense white, somewhat bluish in tint.

[1] Wood, *Phil. Mag.*, 1903, vi. 257.

[2] Lehmann, "A Filter for Ultra-Violet Rays, and its Uses," *Verhandl. d. Deutschen Physik. Gesellschaft*, 1910, No. 21; *Pflügers Archiv f. Physiologie*, cxlii.

Mucous membrane of small intestine: light green.

Spleen: dark brown.

Liver: dark brown, somewhat greenish.

Kidney: dark yellow green.

Section of lung: dark brown, somewhat greenish; transverse sections of the larger vessels of bronchi fluoresce light blue.

Aorta: intense light yellow.

Muscles: intense light green.

Fat: light green.

Bones: light blue.

Cornea and lens of the eye: intense light blue.

J. E. B.

REFERENCES

Culmann, P., "La Microphotographie au moyen des radiations ultra-violettes," *Bull. de la Soc. française de Photographie*, IIe série, tome 22, No. 3.

Dieck, W., "Mikrophotographische Aufnahmen mit ultravioletten Strahlen," *Deutschen Monat. f. Zahnheilkunde*, 24 Jahr. Jan.; "Das Photomikroskop für ultraviolette Strahlen und seine Bedeutung für die histologische Untersuchung, insbesondere der Hartgewebe," *Sitzungs. der Gesellsch. naturf. Freunde*, Jahr 1906, No. 4.

Ernst, Harold C., and Wolbach, S. B., "Ultra-Violet Photomicrography (a Preliminary Communication)," *Journ. of Med. Res.* (n.s. ix.), April 1906, xiv. No. 3.

Franz, V., "Photographieen mit ultraviolettem Lichte," Teil i., *Vom Ovarialei der Knochenfische*; *Zeitschr. f. wiss. Mikr. und für mikros. Tech.* Band 27; "Photographieen mit ultraviolettem Lichte," Teile ii. u. iii., *Vom Vogelauge*; *Archiv für vergleich. Ophthal.*, Juli 1910, Band i. Heft 3.

Hertel, E., "Experimentelles über ultraviolettes Licht." *Bericht über die 31. Versamml. der Ophth. Gesellsch.*, 1903, Heidelberg.

Köhler, A., "Mikrophotographische Untersuchungen mit ultraviolettem Lichte," *Zeitschr. für wiss. Mikros. und für mikros. Technik*, 21; "Eine mikrophotographische Einrichtung für ultraviolettes Licht und damit angestellte Untersuchungen organischer Gewebe," *Physik. Zeitschr.*, 5. Jahr. No. 21.

Köhler, "La Microphotographie en lumière ultra-violette," *Revue gén. des sciences pures et appliquées*, 28 février 1905, 16e année, No. 4; "Ultra-violet Light in Microscopy," *The Opt. Instru. Monthly*, June 1905, i. No. 1.

Köhler, A., u. von Rohr, M., "Eine mikrophotographische Einrichtung für ultra violettes Licht," *Zeitschr. für Instrumen.*, 12 Dezember, Jahr 1904.

Köhler, A., "Swingles Einstellverfahren für die Mikrophotographie mit ultra-violettem Lichte," *Zeitschr. f. wiss. Mikros. und f. mikros. Tech.* 24; "Aufnahmen von Diatomeen mit ultraviolettem Lichte," *Jahrb. für Photographie und Reproduktions* für das Jahr 1909.

Kruis, L., "Mikrofotografické zobrazovaní ultraviolvým světtem," *Živa*, Band 16, 6.

Latour, B., "Le Microscope nouveau de Zeiss à lumière ultra-violette." *Cosmos*, 51e année, No. 1102.

Mouton, H., "Dispositiv microscopique pour l'emploi de la lumière ultraviolette (λ=275 mm.) et examen des tissus organiques faitesavec ce dispositiv," *Le Radium*, 2e année, No. 2.

De Orueta, Domingo D., *Nota sobre la luz ultravioleta y sus applicaciones al microscopio* (Madrid), 1912, Imprenta Alemana Fuencarral.

Rijkens, R. G., "De Microfotografie met ultraviolet-licht," *De Natuur*, 25 Jahr., 15 Januari 1905.

Sabine, W. C., "The Optical Advantages of the Ultra-violet Microscope," *Journ. of Med. Res.* (n.s. ix.) xiv. No. 3.

v. Schrötter, H., "Beitrag zur Mikrophotographie mit ultraviolettem Lichte nach Köhler," *Virch. Arch. für pathol. Anat. und Physiol. und f. klin. Med.*, 1906, clxxxiii. 343.

Stempell, W., "Über die Auflösung feinster organischer Strukturen durch Mikrophotographie mit ultraviolettem Lichte," *Zeitschr. f. physik. Chemie*, lxvii. 2; "Über Nosem bombycis Nägeli nebst Bemerkungen über die Mikrophotographie mit gewöhnlichem und ultraviolettem Lichte," *Arch. f. Protistenkunde*, 1909, xvi. 281.

Swingle, Walter T., and Briggs, Lyman, J., "Improvements in the Ultra-violet Microscope," *Science*, new series, xxvi. No. 658.

Vianna de Lima, A., "Les limites théoriques et pratiques du pouvoir résolvant et le microscope nouveau de Zeiss pour l'emploi de la lumière ultra-violette," *Rev. des Sciences photographiques*, 2e année, No. 9.

MILLILAMBERT: one-thousandth part of a Lambert (*q.v.*). It is a unit of brightness equal to $10/\pi$ candles per square metre—*i.e.* to 0·92 equivalent foot-candles. See "Photometry and Illumination," § (2).

MILLIPHOT: one-thousandth part of a phot (*q.v.*), and therefore equal to 10 lux or 0·92 foot-candles. See "Photometry and Illumination," § (2).

MINE GASES, DETECTION OF, BY BEATS. See "Sound," § (53) (ii.).

MIRRORS, HALF-SILVERED. See "Silvered Mirrors and Silvering," § (5).

MIXTURE CURVES: the curves showing the proportion of the three primaries of a trichromatic colorimeter required to reproduce the colours of the spectrum. See "Eye," § (10).

MODULATION: a term used in music to denote change of key or mode in the course of a piece of music. See "Sound," § (4).

MONOCHROMATIC ILLUMINATOR: an instrument for separating the constituents of a white or complex light and illuminating an object with light of one wave-length only. See "Spectroscopes and Refractometers," § (21).

MOSELEY'S LAW OF ATOMIC STRUCTURE. The positive charge on the nucleus of an atom is N units, where N is the atomic number, *i.e.* the sequence number of the element in the periodic table, and there are N electrons, each of unit negative charge, surrounding it, to counterbalance the nucleus and form the atom. See "Crystallography," § (20).

MOTOR CAR HEADLIGHTS, PHOTOMETRY OF. See "Photometry and Illumination," § (113).

— N —

Nakamura's Polarimeter. See "Polarimetry," § (11) (ii.).

Nautical Astronomy. See "Navigation and Navigational Instruments," § (18).

NAVIGATION AND NAVIGATIONAL INSTRUMENTS

I. Introduction

§ (1) Methods employed.—The principal demand that the navigator of a ship makes upon physical science is for instruments that will enable him to determine his position upon the surface of the earth, and for instruments that will enable him to determine the direction in which his ship is moving.

Until a recent date methods of navigation could be classified either under the heading of "dead reckoning" or else of nautical astronomy. In addition, methods of determining position from the bearings of recognisable landmarks, and of identification of position from a series of soundings, occupy a subsidiary place in the science of navigation. Recently an entirely new method of position fixing has been developed, namely directional wireless telegraphy, a method which has such great possibilities that it is conceivable that in time it will entirely displace astronomical methods.

§ (2) Leading Marks.—The simplest of all methods of position fixing is by leading marks when in sight of a coast. A shoal can be marked and avoided by beacons set up on shore or by natural landmarks noted as being in line when just clear of the danger. By watching the relative changes of bearing of these marks the navigator keeps his ship in safe waters. An extension of this method (which requires the use of four shore objects) is to utilise three only and the angles they subtend at the ship. If three points, A, B, and C, can be identified on the chart and if the angular distance from A to B is α and the distance from B to C is β there is a unique position determined by the point of intersection of two circles. The instruments required for such position fixing are a sextant to measure the two angles α and β and, to avoid the trouble of geometrical construction, a station pointer to lay off the position on the chart.

Another method is to take the bearings of two known objects by means of the azimuth compass. Lines of bearing in the reversed directions from those observed are then drawn through the objects upon the chart and their point of intersection gives the position of the ship. With one object only the fix can only be made by taking two bearings at different times and laying down upon the chart the "run" of the ship in the interval. For such a fix the navigator must have means of determining the direction and speed of the ship through the water, and these data, combined with the observed leeway and the set and drift of the tides as given by tide tables, enable him to complete the fixing of his position. Speed is measured by log line, which in its primitive form was simply a log of wood dropped over the stern and line paid out as fast as it would run, speed being measured by the number of knots of the line that ran out in a specified time, usually fourteen or twenty-eight seconds. The modern form of log indicates continuously the distance a ship has run. For rough purposes the rapidity of revolution of the propeller is also a measure of the speed.

§ (3) Astronomical Methods.—Such methods as have been described above are applicable only when the ship is in sight of land. For cross-seas navigation the position is known provided the course, *i.e.* the direction, and speed are continuously and accurately known. But an error in the determination of the course or an error in the estimation of the speed will have a cumulative effect upon the ship's position, so that it is clear that dead-reckoning methods cannot be relied on to give accurate information to the navigator for any length of time. Besides the inaccuracies of the log and the inaccuracies of steering or of the compass itself, there are also unknown effects from wind and from ocean currents. From time to time, therefore, it becomes necessary to have recourse to astronomical methods in order to fix the position and so obtain a new and accurate point of departure from which a new dead reckoning can be made. Ordinarily the navigator likes to have an astronomical fix at least once a day; in a ship running at high speed sights are taken more often still when the weather permits. With an overcast sky nothing, until the advent of directional wireless telegraphy, was available to the navigator but dead-reckoning methods.

Every astronomical observation for position is a measurement of the angular altitude of a heavenly body above the horizon at a noted time. The altitude is measured by a sextant and the time obtained from a chronometer whose error on Greenwich Mean Time is known with sufficient accuracy. Formerly this error used to be obtained from the known error when the ship was last in port, and from the rate determined from a previous error. Nowadays a ship with wireless can obtain

the error of her chronometers by time signal in almost all parts of the world.

§ (4) THE SUMNER LINE.—A sight having been taken, the observed altitude and time give the navigator a means of drawing upon his chart a certain line of position or "Sumner Line." The position of this line can be realised from *Fig.* 1. Let X be the subsolar (substellar

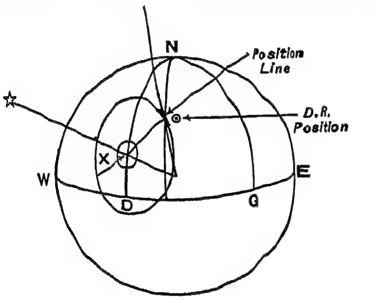


FIG. 1.

or sublunar) point, that is, the one point upon the earth's surface at which the sun's centre is at the zenith at the time the sight is taken. The earth being treated as spherical, the subsolar point is that place at which a line drawn from the earth's centre to the sun's centre will cut the earth's surface. Such a line being a radius of the earth is a vertical line at this point, and the sun, at that point and at that time, has an altitude of 90° above the horizon. The latitude, XD, of the subsolar point is the sun's declination and the longitude, GD, is the Greenwich apparent time converted into angle. For the case of a star the latitude is the declination and the longitude the star's Greenwich hour angle at the time.

If a circle of 60 sea miles radius be described round the subsolar point, its circumference contains all those points on the surface of the earth whose verticals are inclined at an angle of one degree with the vertical of X. Hence at the instant in question this circle is the locus of all points for which the altitude of the sun is 89°. If the circle has 120 sea miles radius it becomes the locus of points for which the altitude is 88° and so on. The curves of constant altitude are in fact a series of small circles drawn round the subsolar point as pole and spaced in exactly the same way as the parallels of latitude round the pole of the earth. The deductions that can therefore be made from an ordinary navigational "sight" are simply that the ship's position is somewhere on a certain small circle that can be drawn, if need be, for its pole and its radius are known.

As the navigator always knows the position of his ship with approximate accuracy, he is concerned with a small portion only of this small circle, namely that part of its circumference which is in the neighbourhood of his dead-reckoning position. For such length of the circumference as he usually has to consider it is sufficient for him to treat the curve when projected on to the Mercator chart as being straight, and such a line is called a position line or Sumner line, after Captain Sumner, an American merchant-service captain who first called attention to the use of them.

The methods by which the position of the Sumner line can be determined are various, but the method of false positions devised by Marc St. Hilaire is the one now most commonly practised. A position is adopted which is known to be not far different from the ship's actual position, and the altitude of the sun or other heavenly body is calculated for that spot and for the instant at which the sight was actually taken. A comparison is made between this calculated altitude and that measured by the sextant, and the difference between the two, measured in minutes of arc, represents the distance, measured in sea miles, separating the assumed position from the Sumner line. If the observed altitude is the greater, then the Sumner line is nearer to the subsolar point than the assumed position and *vice versa.* The Sumner line is always drawn at right angles to the true bearing or azimuth of the heavenly body observed, so that the distance of the assumed position from the Sumner line is always measured in the direction of the true bearing.

§ (5) SOURCES OF ERROR. REFRACTION.—As in all instruments for precise measurement, one of the main considerations in the design is the errors that are likely to arise and to be present in the measured values. The effect of an error in the measurement is very apparent, as each minute of error moves the Sumner line one sea mile parallel to itself. The accuracy with which the position line can be laid down upon the chart is precisely the accuracy (apart from errors of drawing) with which the sun's altitude can be measured. Errors enter into the final result either from errors of the sextant or inaccuracies due to external causes, of which the only one which needs consideration is the variable atmospheric refraction.[1] This will affect both the position of the sun in the sky and the apparent position of the horizon, and it is obvious that the latter will be the more displaced of the two. The depression of the visible horizon in minutes of arc, if the atmosphere had no refractive effect, is $1 \cdot 063 \sqrt{h}$, where h is the height of the observer's eye in feet. Atmospheric refraction reduces this amount by approximately 8 per cent, and the tables for dip, as the depression of the horizon is also called, are calculated

[1] See "Trigonometrical Heights," etc., Vol. III.

from the formula $0{\cdot}98\sqrt{h}$. For abnormal states of the barometer and temperature correcting tables are given, and in the majority of cases the results are sufficiently close for the ship's navigator. Circumstances may, however, arise in which these dip tables are widely erroneous. The most likely conditions for error are a calm clear day with a warm air and a colder sea. There will then be an inversion of temperature in the lower layers of the atmosphere, and the refraction factor in the dip being based on the supposition that the atmosphere is in a state of equilibrium in which the temperature falls uniformly at the rate of 1° C. per 600 feet is in these circumstances quite false. In special cases errors in the dip tables have been known to be so great that the sea horizon is refracted above the true horizontal and the sea as seen from the deck of the ship appears saucer shaped. Errors of four or five minutes under the conditions quoted above appear to be common; several reliable reports have been made of the dip being as much as twenty minutes in error.

When, as in cable laying or in marine surveying, accurate position fixing is needed, it is customary to combat errors of refraction by taking sights in a series of azimuths and determining the position from the Sumner lines so obtained. Sometimes subsidiary instruments are used to measure the dip at the time of the observation. These are described later.

Errors in refraction do not affect the heavenly body observed to the same extent unless its altitude is very low. Usually if the altitude is not less than 20° any inversion of temperature in the lower levels of the atmosphere will hardly affect the apparent position of the body in the sky, and the tabulated values of the refraction in altitude will be good enough.

§ (6) Other Sources of Error.—Other errors which affect the accuracy of the position line are those of the sextant itself, namely the first order zero error and the second order perpendicular error, side error, and collimation error. These are explained in the description of the sextant.

The possibility of the navigators being able to take any sights at all depends upon the sea horizon being visible. Frequently it is obscured by low-lying banks of mist. At night-time it is always difficult to distinguish it except in bright moonlight. When a telescope is used with the sextant the difficulty is greater still. Only a portion of the exit pupil of the telescope is occupied by light from the horizon, and there is always loss by internal reflection at the various glass air surfaces of the telescope. Taking both these into consideration, it is found that the intrinsic brightness of the horizon is less than half when viewed through the telescope of what it is when viewed by the unobstructed naked eye. Through such reasons as these it has become a very common practice of navigators to limit their star observations to morning and evening twilight during the half-hour or so when the horizon is still reasonably clear and the brighter stars can be picked up without difficulty. One direction in which many attempts to solve these difficulties of the navigator have been made is the provision of something which will replace the sea horizon as a basis of measurement of altitude. All these attempts have been made with the idea of providing some sort of level or vertical controlled by gravity. The reason for their failure is referred to in the description of the sextant.

§ (7) Fixing Position.—From what has been stated it will be seen that observations of two heavenly bodies, or two observations of the same body separated by a distinct lapse of time, are necessary for the determination of a fix. During the daytime the second course has to be adopted. It is true that occasionally the moon can be used, but for some reason, possibly because the corrections are troublesome, few navigators make use of it. The planet Venus, when at her greatest distance from the sun, can also be used at times, but it needs a very skilled navigator to pick the star up in broad daylight and to measure her altitude.

Attempts have also been made to obtain a bearing of the sun with sufficient accuracy to enable the position of the ship on the Sumner circle to be identified. If the bearing can be taken with the certainty that the error does not exceed a quarter of a degree, then the ship's position on the Sumner circle can be limited to a length of a few miles, the amount depending upon the altitude. But it is difficult to get a magnetic or gyro compass to maintain such accuracy, and the error cannot be measured to that precision at sea except by Polaris, the only star whose bearing can be sufficiently accurately calculated when the ship's position is only roughly known.

Another direction in which improved aids to navigation are likely to be forthcoming is from the extensive experiments in sound ranging that have been carried out in the recent war. These experiments have covered a very wide field and have achieved many valuable successes. Among these may be instanced directional hydrophones, by means of which distant submarine signals can be located in direction at ranges up to fifty miles. For approaching a coast in foggy weather the peace time uses of such devices are obvious. Again, the experience of war has shown that it is possible, by exploding a charge under water, to time the echo reflected back from

another ship with such a nice degree of accuracy that her range can be calculated from it. It is likely that such research will make it possible to take deep-sea soundings with infinitely less trouble than was attendant upon the old methods.

II. Instruments used in Navigation

§ (8) Methods of Dead Reckoning.—When the navigator has determined his position (latitude and longitude) at a certain instant, and subsequently knows his courses, speeds, and lengths of time running on those courses, he has sufficient information to supply him with the position at any time afterwards. The new position can be determined either by plotting the track upon the chart or by working out a "traverse," which means that each element of the distance run must be resolved into its north-south and east-west components. The algebraic sum of the former gives the "diff. lat." from his previous position, the latter his "departure" from the same spot. The "departure" being found, the difference in longitude is deduced from the relation

Diff. long. = Departure × secant mid latitude.

The whole of the resolution of the different elements of the run into components can be done by the traverse table.

The accuracy of the final estimation of position depends upon the accuracy of the arguments—course and distance—with which the traverse table is entered. "Course" is rendered inaccurate by unsteady steering, and by erroneous estimation of the leeway made. "Speed" has errors due to inaccuracies of the log.

The two instruments concerned with dead-reckoning estimation of position are the Compass (magnetic or gyrostatic) and the Log line.

§ (9) The Magnetic Compass. (i.) *Description.*—The magnetic compass is undoubtedly the most important instrument of navigation in use in the ship. It consists of a circular card, divided into degrees and "points," to which is attached a series of magnetic needles. The whole, balanced upon a central pivot and kept in a horizontal position through the weight being below the point of support, turns under the earth's magnetic force so that the needles lie in the magnetic meridian. The reading of the card against a "lubber-line" in the bowl then indicates the direction of the ship's head.

Compasses are either of the liquid type or else of the dry-card type. In the former case the whole bowl is filled with a mixture of alcohol and water, the formula for naval compasses being ordinarily 67 per cent distilled water and 33 per cent pure alcohol. For air purposes, where much lower temperatures are encountered, pure alcohol is used. The object of the liquid is to increase the damping effect and cause the compass to settle down more quickly if by any chance it is disturbed.

The support of the card is usually in the form of a sapphire cap resting on an agate point.

The needles in most cases are disposed symmetrically in two parallel bundles on either side of the centre and at such distances as will give the whole equal moments of inertia about any diameter of the card. Unless this equality exists the principal moments of inertia will lie, from considerations of symmetry, in the north-south and east-west lines. If such a compass is subject on board to a forced oscillation about any other diameter of the card than a cardinal one, an oscillation of the card in its own plane will be set up with a period which increases as the difference between the two principal moments of inertia is lessened. As forced oscillations of this description are very likely to result from the vibrations of the main engines or of other machinery, the equality of the two principal moments of inertia is of considerable importance. With two single needles the distance apart should be equal to the length of the needles divided by $\sqrt{3}$.

The relative values of the strength of the horizontal component of the earth's magnetic field, the viscosity of the liquid, friction of the pivot, etc., and the moment of inertia about the axis determine the period at which the card will oscillate about its equilibrium position and the rate of decay of the amplitude of this oscillation. Usually with a liquid compass the period for the latitude of Greenwich is from 10 to 18 seconds from rest to rest and the decay factor, *i.e.* the ratio of a half-swing from rest to rest to the one preceding it, about one-half for each successive half-swing.

In certain compasses used in aircraft the damping has been increased and the moment of inertia reduced to such an extent that the compass is aperiodic and goes directly back to the meridian without oscillation.

The direction of the ship's head, as shown by the compass, requires correction for variation and deviation. The former term denotes the angle between the magnetic and the geographical meridians and is obtainable always from the Admiralty charts, in which the amount at the date of publication and the annual change are indicated. The deviation of the compass is the angular deflection of the needle from the magnetic meridian due to the influence of the permanent and induced magnetism of the iron and steel used in the construction of the ship.

(ii.) *Historical.*—The origin of the magnetic compass is lost in obscurity. It is apparently referred to in Chinese and Siamese manuscripts

of the fourth century. Possibly it antedates the Christian era. Unquestionably it remained a very crude instrument indeed until the fourteenth century, when the Neapolitan Flavio Gioia made very considerable improvements in it, but even then it was so very approximate in its indication of the north point that it was always believed to indicate the true north, and it was not until 1555, when Martin Cortez of Seville published a book on navigation, that any mention is made of magnetic variation. Twenty years later magnetic dip was discovered by Norman, and from data supplied by seamen curves of magnetic variation and dip were gradually constructed. Towards the end of the seventeenth century differences in the variation as supplied by different ships began to be manifest. Denis in 1666, and Dampier in 1691, instanced anomalies of a similar kind, and Wales, who was astronomer to Cook's expedition to the South Seas, first asserted that the direction of the ship's head affected the compass. In 1794 Mr. Downie, master of H.M.S. *Glory*, observed that "the quantity and vicinity of iron in most ships has an effect in attracting the needle." In 1801 Captain Flinders, whose name is still preserved in connection with the "Flinders bar," found that a change in the position of the binnacle affected the compass, and as a result of his representations to the Admiralty a series of experiments was made at Sheerness on several of H.M. ships, but the scope of the work was too limited to permit of really general deductions being drawn. Professor Barlow, after an extensive course of experiments, proposed in 1819 to counteract the effect of large masses of iron used in the construction of the ship by a circular plate suitably placed in proximity to the compass, and in 1835 Captain Johnson, the Admiralty Superintendent of Compasses, found during an investigation into the applicability of Barlow's plate to the compasses of the iron ship *Garryowen*, that the ship herself was permanently magnetic and that the plate failed to correct the resulting compass error. As a result of Johnson's experiments the Admiralty commissioned Airy, then Astronomer-Royal, to make further experiments on board the iron vessels *Rainbow* and *Ironsides*, and his conclusions from these experiments were that the permanent magnetism of the ship could be counteracted by the location near the compass of suitably placed permanent magnets.

While these investigations were being made in England, Poisson was also engaged in similar work in France. His work, which was of a more academic character, was presented to the Academie des Sciences in 1824. This celebrated memoir on the Theory of Magnetism is generally recognised as forming the foundation of the Theory of the Deviations of the Compass. In 1838 he contributed a second memoir in which, although recognising the possibility of permanent magnetism being present, he yet presumed that it was of very small amount, and omitted it in his solutions.

Unlike Poisson, Airy started with the recognised necessity of dealing with the permanent magnetism, but the deductions he made from the *Ironsides* experiments were not fully justified, and subsequently a serious divergence of opinion arose between Airy and Scoresby respecting the validity of the former's conclusions. The main point at issue was whether the rules devised by Airy for the correction of the compass applied only in the place at which the corrections were made or whether they could also be considered reasonably accurate for changes of latitude. Scoresby also affirmed that the so-called permanent magnetism of the ship, produced by hammering, etc., during construction, was only sub-permanent; that a large part of it disappeared during the early voyages; and that it might be very largely changed by the buffeting that a ship might receive if she continued on the same course for any length of time in heavy weather. The loss of an important ship, the *Taylor*, on her first voyage, in 1854, a loss which was in some quarters attributed to the defects of Airy's conclusions, led to the formation of the very important Liverpool Compass Committee, which issued three reports on the Deviations of the Compass —in 1856, 1857, and 1861. Following these reports the Admiralty ordered further investigation, and a mathematical examination of the subject was carried out by Archibald Smith, very largely based on Poisson's original work. Subsequently, under the direction of Captain F. J. Evans, R.N., Superintendent of the Admiralty Compass Department, and of Archibald Smith, the *Admiralty Manual of the Deviations of the Compass* was published in 1862, and passed through successive editions in 1863, 1869, 1874, 1882, 1893, 1901, and 1912.

The Admiralty Manual, with its investigation and rules for correction, is now looked upon as the standard work on the subject, and has been adopted as the basis of similar publications in other countries.

Very few changes were necessary in the first five editions, but in the sixth, brought out in 1893, certain modifications of detail had to be introduced on account of the greater use of steel in ship construction; but in principle there is little change since the work first appeared, and the only controversial points that have aroused any discussion have been such matters as the comparative merits of the liquid as against the "dry-card"

compass, the construction of the cap and pivot, and the most suitable period of oscillation.

With the advent of aeroplanes and airships the need of compasses for their special use became apparent, and experimental work is still in progress for the purpose of evolving the best type. One of the principal difficulties in this connection lies in the fact that an aeroplane in turning is subject to very large horizontal acceleration, and the compass card "banks" in consequence at angles of possibly 60° or 70° with the horizontal. Under such circumstances the vertical component of the earth's magnetism may have as much or even more effect on the needle than the horizontal. In particular it is quite possible, when turning away from north, for the compass to go off in the wrong direction. When in sight of the ground this is comparatively unimportant, but when flying in cloud the unreliability of the compass in indicating turn can easily result in the pilot's getting his machine into a "spinning nose dive." It appears impossible to provide a compass that will fulfil all the requirements of the air pilot, and up to the present the "aperiodic" previously referred to seems to be the best compromise.

§ (10) MATHEMATICAL THEORY OF THE COMPASS. (i.) *Deviation Error.* — The mathematical theory of the deviations of the compass has to take account of the permanent magnetism of the ship, of the induced magnetism of soft iron changing with each change of course, of the varying values of the earth's magnetic field in different latitudes, and to include the effects that are due to the rolling, pitching, and yawing that the ship may be subject to in a seaway. Of these three types of motion that of rolling is the most important. In the days of sailing ships the vessel would be heeled over to leeward for long periods, and the "heeling error" was of primary importance. In a steamship rolling takes place about the upright position, and consequent deviations due to heeling are of alternate sign, so that their relative importance is less than it was in sailing-ship days. As the whole of the investigation of heeling error has been made on a statical basis, it is quite probable that the formulae obtained are not true for a ship with a considerable degree of roll when dynamical considerations enter into the question.

For the purpose of considering the quantitative values of the deviation, take as axes of reference Ox in the fore and aft line (ahead) through the compass, Oy in a horizontal athwartship direction (to starboard), and Oz in a vertical direction downwards, the assumption being that the ship is for the present on an even keel (see *Fig.* 2).

Let X, Y, and Z be the components of the earth's field at the origin, and P, Q, and R the components of the ship's field (permanent magnetism).

Then the total field will have components

$$X' = X + aX + bY + cZ + P,$$

$$Y' = Y + dX + eY + fZ + Q,$$

$$Z' = Z + gX + hY + kZ + R,$$

where $a, b, c, d, e, f, g, h, k$ are parameters dependent upon the distribution of soft iron in the ship. These equations are due to Poisson.

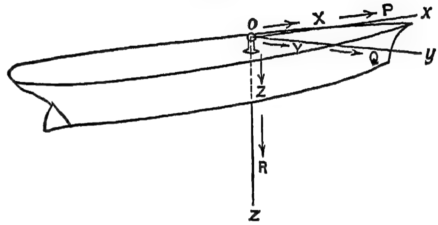


FIG. 2.

If H and H′ are the horizontal force of the earth and the horizontal force of the earth and ship respectively, θ the dip, ζ the magnetic course, ζ'

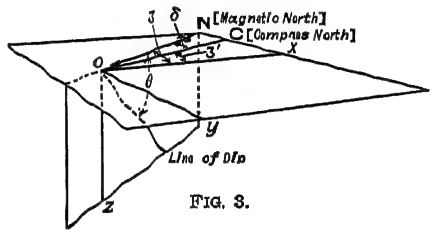


FIG. 3.

the compass course, and $\delta\ (=\zeta-\zeta')$ the deviation, we have (see *Fig.* 3)

$$X = H\cos\zeta, \quad Y = -H\sin\zeta, \quad Z = H\tan\theta,$$

$$X' = H'\cos\zeta', \quad Y' = -H'\sin\zeta';$$

by substitution and reduction we obtain

$$\frac{H'}{H}\sin\delta = \frac{d-b}{2} + \left(c\tan\theta + \frac{P}{H}\right)\sin\zeta$$

$$+ \left(f\tan\theta + \frac{Q}{H}\right)\cos\zeta + \frac{a-e}{2}\sin 2\zeta + \frac{d+b}{2}\cos 2\zeta,$$

and

$$\frac{H'}{H}\cos\delta = 1 + \frac{a+e}{2} + \left(c\tan\theta + \frac{P}{H}\right)\cos\zeta$$

$$-\left(f\tan\theta + \frac{Q}{H}\right)\sin\zeta + \frac{a-e}{2}\cos 2\zeta - \frac{d+b}{2}\sin 2\zeta.$$

The former of these equations gives the strength of the couple tending to turn the compass needle eastwards, the latter the strength of the couple tending to turn to north. The mean value of the former is $(d-b)/2$, which is usually, but not necessarily, zero; the mean value of the other is $1+(a+e)/2$. This we denote by λ. In iron vessels λ is nearly always less than unity. The quantity λH, which

is the mean value of $H' \cos \delta$, or of the component of the earth and ship directed to magnetic north, is usually known as "mean force to north."

If we write

$$\frac{1}{\lambda}\frac{d-b}{2}=\mathfrak{A},\quad \frac{1}{\lambda}\frac{a-e}{2}=\mathfrak{D},\quad \frac{1}{\lambda}\frac{d+b}{2}=\mathfrak{E},$$

$$\frac{1}{\lambda}\left(c\tan\theta+\frac{P}{H}\right)=\mathfrak{B},\quad \frac{1}{\lambda}\left(f\tan\theta+\frac{Q}{H}\right)=\mathfrak{C},$$

we obtain

$$\tan\delta=\frac{\mathfrak{A}+\mathfrak{B}\sin\zeta+\mathfrak{C}\cos\zeta+\mathfrak{D}\sin 2\zeta+\mathfrak{E}\cos 2\zeta}{1+\mathfrak{B}\cos\zeta-\mathfrak{C}\sin\zeta+\mathfrak{D}\cos 2\zeta-\mathfrak{E}\sin 2\zeta},$$

which gives the deviation on any magnetic course when the coefficients $\mathfrak{A}$, etc., are known.

In terms of the compass course the expression for the deviation is more complicated, but can be written in the form

$$\begin{aligned}\delta=&A+B\sin\zeta'+C\cos\zeta'+D\sin 2\zeta'+E\cos 2\zeta'\\&+F\sin 3\zeta'+G\cos 3\zeta'+H\sin 4\zeta'+K\cos 4\zeta'\\&+L\sin 5\zeta'+M\cos 5\zeta'+N\sin 6\zeta'+\ldots\end{aligned}$$

in which A, B, C ... are derivable from $\mathfrak{A}$, $\mathfrak{B}$ ... In view of the fact that $\mathfrak{B}$, $\mathfrak{C}$, $\mathfrak{D}$ are small quantities of the first order, $\mathfrak{A}$ and $\mathfrak{E}$ small quantities of the second order, and if we go to terms of the third order, we find

$$A=\mathfrak{A}$$

$$B=\mathfrak{B}\left(1-\frac{\mathfrak{D}}{2}+\frac{\mathfrak{B}^2}{8}+\frac{\mathfrak{C}^2}{8}+\frac{\mathfrak{D}^2}{4}\right)-\frac{\mathfrak{E}\mathfrak{C}}{2},$$

$$C=\mathfrak{C}\left(1+\frac{\mathfrak{D}}{2}+\frac{\mathfrak{B}^2}{8}+\frac{\mathfrak{C}^2}{8}+\frac{\mathfrak{D}^2}{4}\right)-\frac{\mathfrak{E}\mathfrak{B}}{2},$$

$$D=\mathfrak{D},$$

$$E=\mathfrak{E}+\mathfrak{A}\mathfrak{D},$$

$$F=\frac{\mathfrak{B}\mathfrak{D}-\mathfrak{C}\mathfrak{E}}{2}-\frac{\mathfrak{B}^3}{24}+\frac{\mathfrak{B}\mathfrak{C}^2}{8}-\frac{3\mathfrak{B}\mathfrak{D}^2}{8},$$

$$G=\frac{\mathfrak{C}\mathfrak{D}+\mathfrak{B}\mathfrak{E}}{2}+\frac{\mathfrak{C}^3}{24}-\frac{\mathfrak{C}\mathfrak{B}^2}{8}+\frac{3\mathfrak{C}\mathfrak{D}^2}{8},$$

$$H=\frac{\mathfrak{D}^2}{2},$$

$$K=\mathfrak{D}\mathfrak{E},$$

$$L=\frac{3\mathfrak{B}\mathfrak{D}^2}{8},$$

$$M=\frac{3\mathfrak{C}\mathfrak{D}^2}{8},$$

$$N=\tfrac{1}{3}\mathfrak{D}^3.$$

Succeeding coefficients would all be of fourth or higher orders. Similar expressions for A_1, B_1, C_1, etc., when the expansion is made in terms of the *magnetic* course can also be obtained.

In actual practice it is not necessary to go further in the expansion than the first five terms, and provided the value of δ does not exceed about 20° we can write

$$\delta=A+B\sin\zeta'+C\cos\zeta'+D\sin 2\zeta'+E\cos 2\zeta'.$$

The deviation then consists of a constant part A, a semicircular part $B\sin\zeta'+C\cos\zeta'$, and a quadrantal part $D\sin 2\zeta'+E\cos 2\zeta'$. The constant part A is always small for a compass placed on the fore and aft line of the ship provided the masses of soft iron are symmetrically placed, as most usually happens. The semicircular deviation is due partly to the permanent magnetism of the ship and partly due to induced magnetism; the quadrantal deviation to induced magnetism only. The correction of the former is made by suitably placed permanent magnets and by the "Flinders Bar"; of the latter by suitably placed masses of soft iron, usually two spheres placed one on each side of the binnacle. The next term of the expansion, the sextantal deviation, represents the effect of magnetism induced in the soft iron (in effect, in the correcting spheres) by the magnets of the compass itself. In the particular case of spheres symmetrically disposed, the sextantal term becomes zero provided the distance between the two needles is equal to the length divided by $\sqrt{3}$, a condition which fortunately is satisfied owing to the need of the card to have equal moments of inertia.

It appears, then, that after the compass has been approximately corrected by permanent magnets and spheres the approximate value for the outstanding deviation is that given just above.

(ii.) *Heeling Error.*—When the ship is inclined from her normal position, modifications in the formula for the deviation have to be made. If i be the angle of heel to starboard (see *Fig.* 4) the new deviation δ_i can be shown

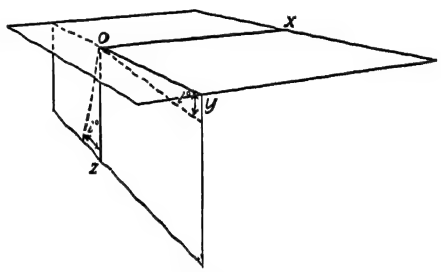


FIG. 4.

to exceed the deviation δ in the upright position by a quantity consisting of a constant part, a semicircular part, and a quadrantal part. When i is small, and when the soft iron of the ship is symmetrically placed, this relationship becomes

$$\delta_i=\delta+\frac{c-g}{2\lambda}i+Ji\cos\zeta'-\frac{c+g}{2\lambda}i\cos 2\zeta',$$

where $$J=\frac{1}{\lambda}\left(e-k-\frac{R}{Z}\right)\tan\theta.$$

The constant part $i(c-g)/2\lambda$, and the quadrantal part $-i(c+g)\cos 2\zeta'/2\lambda$, when taken together are equivalent to

$$\frac{c}{\lambda}i\sin^2\zeta' - \frac{g}{\lambda}i\cos^2\zeta',$$

which will only be zero for all courses provided $c=o$ and $g=o$. These conditions are not usually obtained, but the values of c/λ and g/λ are ordinarily small, and the error represented by these two terms may be safely neglected.

The semicircular heeling error cannot be so neglected, and the error represented by this term has to be specially dealt with in correcting the compass.

The quantity $-J$ is usually spoken of as heeling coefficient, $-Ji$ as heeling error. The correction is made by vertical magnets placed directly below the compass bowl.

(iii.) *The Six Coefficients.*—Owing to the symmetrical disposition of soft iron relative to the compass, four of the Poisson parameters, viz. b, d, f, and h, become zero, and hence A and E are zero. The coefficients B and C of semicircular deviation then depend on $\mathfrak{B}$, $\mathfrak{C}$, and $\mathfrak{D}$. The value of the last term $\mathfrak{D}$ is in all ordinary cases positive, and can be diminished by soft iron correctors placed on either side of the compass. In a perfectly corrected compass D, and therefore $\mathfrak{D}$, must be zero to eliminate quadrantal deviation, and the semicircular will then only disappear provided $\mathfrak{B}=o$ and $\mathfrak{C}=o$. Both these coefficients are compounded of permanent and induced magnetism effects, as is seen from the equations

$$\mathfrak{B}=\frac{(c\tan\theta+P/H)}{\lambda},$$

$$\mathfrak{C}=\frac{(f\tan\theta+Q/H)}{\lambda},$$

and if perfect correction is required in all parts of the world, it is necessary that c and f should be zero and P and Q should be zero. The parameter f is zero from symmetry and c is made zero by means of the Flinders bar, a vertical bar of soft iron placed before or abaft the compass according to whether the coefficient is negative or positive. The remaining parts of $\mathfrak{B}$ and $\mathfrak{C}$ are corrected by permanent magnets placed respectively fore and aft and athwartships for the purpose of reducing both P and Q to zero.

The heeling coefficient $-J$, which can be written

$$\left(D+\frac{\mu}{\lambda}-1\right)\tan\theta,$$

where $$\mu=1+k+\frac{R}{Z},$$

is zero in a ship for which the quadrantal deviation is corrected provided $\mu=\lambda$. Now $\mu=1+k+R/z$, $\lambda=1+\frac{1}{2}(a+e)$, and since $a=e$ is the condition for correction of quadrantal deviation the above equations reduce to

$$\kappa+\frac{R}{Z}=e,$$

a condition that cannot be universally satisfied unless $\kappa=e$ and $R=o$. A correction is, however, possible in any particular magnetic latitude by vertical magnets placed below the compass.

§ (11) Determination of the Errors of a Compass.—By "swinging" a ship it is possible to determine in several ways the amount of the deviation. One of the simplest of those ways is to turn the ship while making repeated bearing observations of a distant point of land. The ship's head is steadied on each point of the compass while the observation is made and a curve of errors subsequently constructed. The diameter of the turning circle of the ship must be small in comparison with the distance of the landmark observed in order to avoid parallactic errors, and the exact magnetic bearing must be obtained by measurement from the chart, or otherwise, in order to discover any constant error (A) of the compass.

If the ship is swung at sea out of sight of land corresponding observations can be made of the sun, but in this case it is simpler to determine the deviations by noting the time of each observation and thence calculating the true azimuth.

An alternative method when close to land is to set up an azimuth compass on shore at a place free from magnetic disturbance and take reciprocal bearings between the landing compass and the ship's instrument on each point. Differences of bearing then represent the deviations.

If the ship is swung in an upright position no indication is obtained of the value of the heeling error when rolling, but sufficient information is forthcoming if the values of λ and μ can be obtained. The former of these, λ, is the ratio of the mean horizontal directive force to north at the compass to the horizontal directive force on shore. It is determined from the times of oscillation of a horizontal magnet set to vibrate first on shore and then on board in the binnacle after the compass bowl has been removed. If the horizontal force on board were the same in all azimuths, we should have $\lambda=H'/H$, but since H' changes with the line of the ship's head it is necessary to use a correcting factor, giving rise to the general expression

$$\lambda=\frac{H'}{H}\cdot\frac{\cos\delta}{1+\mathfrak{B}\cos\zeta-\mathfrak{C}\sin\zeta+\mathfrak{D}\cos 2\zeta-\mathfrak{E}\sin 2\zeta}.$$

The values of $\mathfrak{B}$, $\mathfrak{C}$, $\mathfrak{D}$, and $\mathfrak{E}$ are obtained from the analysis of the curve of deviations.

The second constant μ is obtained from measurements of the vertical component of magnetism. It is the ratio of the mean value of the vertical force on board to the vertical force on shore. The measurement is made by means of a dip circle, which is a magnet mounted so as to turn about a horizontal axis. By turning the whole instrument in azimuth the needle can be made to hang vertically, and the ratio of the squares of the times of oscillations about its equilibrium position on board and on shore leads to the value of Z'/Z. As in the previous case, the value of Z' changes with the direction of the ship's head, and for precise determination of its value it is necessary to make the observations on a number of different directions. A second method of evaluating μ is by means of the heeling error instrument, in which a magnet can also rotate about a horizontal axis. This magnet is exactly balanced before being magnetised and is gravitationally in neutral equilibrium. Hence, when placed so that the needle is free to move in the plane of the magnetic meridian, it would set itself in the line of the dip. A small aluminium weight can be attached to the needle and adjusted so that the needle takes up a horizontal position. The moment of the weight about the axis thus balances the moment of the vertical force. By adjusting the weight for the instrument firstly on shore and secondly on board in the compass bowl the ratio of Z' to Z can be obtained. The angle of dip can also be determined for this instrument.

Having obtained the values of λ, μ, and θ, and knowing also $\mathfrak{D}$ from the deviation curve, the coefficient of heeling error $(\mathfrak{D} - 1 + \mu/\lambda) \tan \theta$ is found.

§ (12) CORRECTIONS OF THE DEVIATION.—In practice it is desirable that a ship's compass should be as nearly as possible corrected on all courses and in all latitudes. It is not possible to obtain this state of affairs with absolute exactitude, partly from the inherent difficulties of the problem and partly from the fact that the so-called permanent magnetism is only subpermanent, and it is therefore essential that the navigator should make frequent azimuth observations of the sun or a star for the purpose of determining the deviation for the course he is actually steering. The order in which the corrections are carried out is

(i.) Quadrantal deviation.
(ii.) Heeling error.
(iii.) Semicircular deviation.

(i.) *The Quadrantal Deviation.*—This consists almost entirely of the term D sin 2ζ, and the coefficient D in an uncorrected compass (*a*) does not change with change of latitude, (*b*) is always positive, and (*c*) remains constant for a great length of time. The correction of D is performed by means of soft iron spheres placed on the port and starboard sides of the binnacles, experience of previous ships of a similar design suggesting the size of the spheres and their position. The ship is then swung with the spheres in position and the new D determined from the table of deviations so obtained.

If this new D has a positive sign the spheres must be moved inwards or smaller spheres provided; if negative the opposite must be done. The necessary data are supplied in Table IV. of the *Admiralty Manual.*

For compasses out of the fore and aft line or between decks in the neighbourhood of unsymmetrically placed masses of iron, a coefficient E may exist. This can be corrected by moving one sphere ahead and the other astern of the athwartship line.

Before touching the heeling error corrections it is usually desirable to take out all or as much as possible of the soft iron semicircular deviation by means of the Flinders bar. This consists of a portion of $\mathfrak{B}$, viz. $(c/\lambda) \tan \theta$, and a portion of $\mathfrak{C}$, $(f/\lambda) \tan \theta$, the latter being zero through the symmetrical disposition of soft iron. The Flinders bar therefore corrects the coefficient *c*, and the magnitude of *c* in an uncorrected ship is known from previous experience. Should it not be so known it can only be determined from value of the semicircular deviations in widely separated latitudes.

(ii.) *Heeling Error.*—The next error needing correction is the semicircular heeling error, consisting of a part due to soft iron and a part due to permanent magnetism. The latter part is much the more important, and no separate means are usually provided for elimination of the soft iron part. The correction is made by means of vertical magnets below the binnacle adjusted up or down until the vertical force in the binnacle is the same as the vertical force on shore. The ratio $\mu : \lambda$ is thus made unity, and since $-\mathrm{J} = (\mathfrak{D} + \mu/\lambda - 1) \tan \theta$ the heeling error is thereby corrected, but will require a new adjustment of the vertical magnets if the ship makes a big change of latitude.

(iii.) *Semicircular Error.*—Except for constant deviation A, there remains only the permanent magnetism part of the semicircular error. This is neutralised by means of permanent magnets, some fore and aft and some athwartship placed below the binnacle, the numbers and sizes being adjusted until the deviation is zero on each of the four cardinal points.

The constant A $(= (d - b)/2\lambda)$ ought to be zero since by symmetry *d* and *b* are both zero. In actual fact there is usually a small A in an uncorrected ship, and this can be most

easily eliminated by a small displacement of the lubber line.

Having made the above corrections as far as can be done, the ship has to be swung and the values of the deviation (now small) used to compile a deviation table, which is checked from time to time, and the compass readjusted, a thing which is particularly necessary after large change of latitude or after heavy weather while steering for a considerable time on the same course.

§ (13) THE GYRO COMPASS. (i.) *General Principles.*—Gyro compasses in a practical form first appeared some two or three years before the European War when they were put on the market by Anschütz of Germany, and by Sperry in the United States. The underlying principles were the same in both compasses and were the practical development of ideas which had been suggested many years previously by Foucault and by Lord Kelvin. The latter described to the British Association in 1884 a gyroscopic model of a spring balance, which contained gyros instead of springs, and a gyroscopic model of the mariner's compass, in which a gyro took the place of the magnetic needle.

The chief points to notice in the behaviour of a spinning flywheel or gyroscope mounted freely so as to be capable of turning in any direction are: (i.) that it offers considerable resistance to any force tending to deflect the axis of spin from its original direction, and (ii.) that in the case of a force at right angles to the axis, when the axis does yield it moves, not in the direction of the force but at right angles to it. Thus, if the spinning axis is initially horizontal, a vertical force will only result in a change of azimuth; but any couple in the horizontal plane will produce a change of tilt and no horizontal motion. It has been suggested that the first-named property of a gyro, its power of resisting disturbing forces, might be used to produce a compass, or at least a direction indicator: the gyro being set with its axle in some given direction at the start, and maintaining that direction throughout the trip. Unfortunately it has, so far, been impossible to provide a mounting so free from friction that the gyro will not slowly drift away from its original direction. Except for certain aeronautical purposes where accuracy within a couple of degrees is required for periods of only half an hour or so, this type of "compass" has therefore never been of any practical use.

Every gyro compass which has met with any success up to now has been definitely north-seeking, and, if forcibly deflected from the meridian, works its way back by a series of damped oscillations like a magnetic compass, but much more slowly. This is accomplished by taking advantage of the earth's diurnal rotation and causing the gyro to be affected by it. This rotation has an influence on the gyro compass which is analogous to the effect of terrestrial magnetism on a magnetic compass. The earth's rotation may be regarded as an angular movement of the vertical in space, and is made use of by causing the gyro to be sensitive to any change in the position of its axis of spin relatively to the local vertical. The obvious method of doing this is to mount the gyro with its axis horizontal in a pendulous frame. The gyro tends to maintain its direction in space while the vertical shifts; its axis thus in general acquires a tilt from the horizontal position, and the pendulousness of the frame then supplies a couple which causes the gyro to "precess" or move in azimuth. If the axis points north and south, the shift of the vertical does not alter the angle between it and the axis, which still remains horizontal; thus no couple is produced and the axis continues to point north and south. This is the principle of Foucault and of Kelvin, applied by Anschütz, Sperry, and others. It is the only principle on which a successful gyro compass has yet been constructed.

(ii.) *Early Forms of Compass.*—Trials of Anschütz and Sperry compasses were carried out by the British Admiralty about 1910, but it was found that although otherwise fairly satisfactory they both showed large deviations when the ship was rolling, unless at the same time the ship's head was towards one of the four cardinal points. This "Inter-Cardinal Rolling Error," as it is called, was overcome in about a year by both makers. Anschütz redesigned his compass entirely and added two more gyros of the same size as the primary gyro, making three in all; at the same time he adopted an entirely new method of damping invented by Schuler, a German. Sperry, however, adhered to the original form of his compass, but added one small subsidiary gyro called the "Floating Ballistic Gyro" to overcome the deviation formerly caused by rolling.

The increasing use of large masses of iron and of electrical machinery in modern ships—particularly in submarines—had by this time greatly impaired the reliability of the magnetic compass, so that the advent of a gyro compass which would not be affected by stray magnetic fields, and which promised to indicate the true north at all times, was particularly timely. After experience with both Anschütz and Sperry compasses under service conditions, the British Admiralty decided to adopt the latter in the Navy. All submarines and first class ships were accordingly fitted with Sperry compasses between 1911 and the end of the war.

The Patent Office records show that the gyro compass problem attracted the attention of many other inventors, but only three need

be specially mentioned here. In 1917 Professor Perry and S. G. Brown produced a compass for sea trials which contained some novel and useful features. A great advance in eliminating undesirable friction was attained by causing the gyro frame to oscillate in its bearings by an intermittent flow of oil. It was claimed that the rolling error was eliminated by making the gyro and its casing strictly non-pendulous, while producing the required couple by mechanically transferring oil from the low side to the high side whenever the compass was tilted. This transfer was accomplished by an air-blast relay controlled by a pendulum. The damping was effected by the same relay and was similar in principle to that proposed by Schuler for the Anschutz compass. This compass was installed in a British light cruiser, with the result that an order for a further number was placed by the Admiralty. Before these could be delivered the Armistice was declared and, owing to the consequent stoppage of new construction, they have not yet been put into service.

A somewhat similar fate befell a compass with two gyros invented by H. L. Tanner in New York and made by the Sperry Gyroscope Co. It was a radical departure from previous models and was noteworthy for the merit of its electrical equipment. The Armistice spoilt the chance of this compass also, so far as the British Navy was concerned, as it had only been installed in two warships. Both the Perry-Brown and the Tanner compasses have, however, met with considerable success in the mercantile marine. As the Sperry compass, more or less modified as will be described later, is almost universal in the British, American, French, and Italian warships, and has a performance which is fairly representative of gyro compasses in general, the following pages refer more particularly to this apparatus.

§ (14) The Sperry Compass.—The main features of the Sperry compass may be understood from *Fig.* 5.

The gyro is simply a flywheel 12 in. diameter weighing 50 lbs., and spun electrically by 3-phase current at a rate of 8600 r.p.m. It is enclosed in an airtight aluminium case *b*, which is exhausted to a good vacuum to prevent air friction and consequent heating. This case is supported in a vertical ring *d* in which it is free to tilt about the horizontal axis *c*. The ring can turn about its vertical axis ff_1, in an outer vertical ring *g*, and carries the gyro case with it. The bearings *f* and f_1 are merely guide bearings and do not support the weight which is borne by a stranded wire attached to the outer ring. This outer ring, called the Phantom Ring, is power-driven by a small electric motor *m*, controlled by contacts on the inner vertical ring. The arrangements are such that the plane of the phantom ring follows faithfully the plane of the inner vertical ring. The wire *e* therefore never gets twisted, and there is no movement of the bearings *f* and f_1, except for a small oscillatory motion of the phantom. The contacts have no "off" position; the motor is always running in one direction or the other, so the plane of the ring *g* is constantly hunting about a mean position coplanar with the ring *d*, with an amplitude of about a quarter of a degree. This allows the ring *d* to turn about its vertical axis ff_1, almost without friction. One may blow once on the side of the gyro case and cause it to

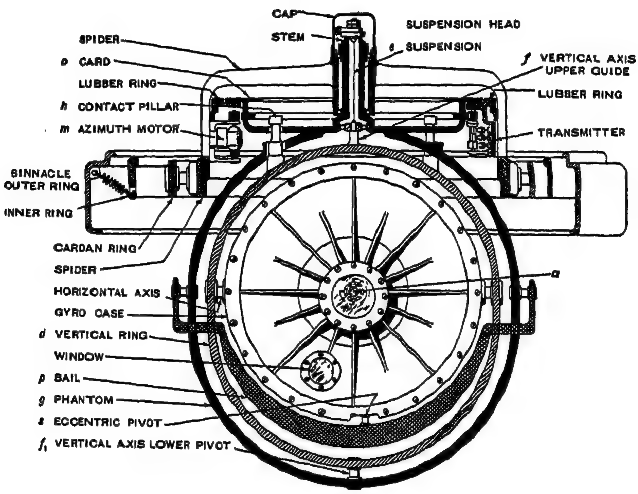


Fig. 5.

make several revolutions if the wheel is not spinning.

The gyro case is as nearly as possible in neutral equilibrium and the necessary pendulous moment is given to it by the separate mass p called the bail, which is hung independently on the phantom ring, and connected to the bottom of the gyro case at the point s. In order to avoid inter-cardinal rolling error, it is arranged that this point shall always be in the same position relative to the true vertical line through the centre of the gyro, even when the gyro case and its supporting rings are being swung east and west by the motion of the ship. The connection at s is therefore made by two rollers on one stem, the upper one sliding in a groove in the case and the lower one sliding in a similar groove in the bail. The roller stem is hung at the bottom of a short pendulum suspended from a. This pendulum, having a quick period, would itself be set swinging by the ship's motion if it were not stabilised by the small auxiliary gyro attached to it and known as the "floating ballistic gyro." The compass card o is carried by the phantom ring, and below it are arranged an automatic device for correcting the speed and damping errors, and an electrical transmitter for transmitting the corrected indications to a number of subsidiary dials in various parts of the ship.

§ (15) Theory of the Sperry Gyro.—The theory of the gyro compass,[1] written by H. Crabtree for Messrs. Elliott Bros.' handbook of the Anschütz compass, has been reprinted in that author's *Spinning Tops and Gyroscopic Motion*. The matter is also dealt with by Klein and Sommerfeld in *Theorie des Kreisels*. Sir George Greenhill devotes a chapter to the Gyro Compass in his *Report on Gyroscopic Theory*. All these writers, however, deal with the original single-gyro Anschütz compass, which is now obsolete. So far as is known, no similar mathematical explanation of the Sperry compass has yet been published, although it is in use to-day in all the principal navies of the world. A brief outline of the mathematics of the Sperry compass is therefore given here.

(i.) *General Equations.*—Let the axis of the gyro make an angle α with the meridian and β with the horizontal. Then if β is small the earth's diurnal rotation of ω radians per second has components in latitude λ.

$\omega \cos\lambda \cos\alpha$ about the gyro axle, . (1)

$\omega \cos\lambda \sin\alpha$ about the gyro tilt axis c, (2)

$\omega \sin\lambda$ about the local vertical. . (3)

The component (1) has no visible effect on the gyro, but if it maintains its direction in space, the axle has apparent motions equal to (2) and (3) relatively to the earth which increase β and α respectively.

As the gyro casing is not in perfectly neutral equilibrium, let C be its pendulous moment; to this is added B the pendulous moment of the bail. These two parts tilt together, so that the torque about the horizontal axis c of the gyro case is $\beta(B+C)$. From the fundamental gyroscopic law[2] this produces a precession in the horizontal plane at the rate $\beta(B+C)/H$, where H is the angular momentum of the flywheel. The apparent precession relative to the meridian is, however,

$$\beta\frac{B+C}{H} - \omega\sin\lambda = -\frac{d\alpha}{dt}. \quad . \quad . \quad (4)$$

Therefore, when the gyro remains at rest in azimuth it must have a tilt whose steady value is $H\omega\sin\lambda/(B+C)$.

In practice it is desirable for the gyro to settle with its axis perfectly horizontal as shown by spirit levels attached to the case. This is arranged by a latitude adjustment which sets the gyro and bail tilt axes out of alignment so that when the gyro axis is level and $\beta=0$, the bail has a tilt $\beta_0 = H\omega\sin\lambda/B$.

The torques about the horizontal axis c are now $C\beta$ due to the case and $B(\beta+\beta_0)$ due to the bail.

Equation (4) may then be written

$$C\beta + B(\beta+\beta_0) - H\omega\sin\lambda = -H\frac{d\alpha}{dt},$$

or since the terms $B\beta_0$ and $H\omega\sin\lambda$ cancel,

$$-H\frac{d\alpha}{dt} = \beta(B+C). \quad . \quad . \quad . \quad (4a)$$

Then
$$-H\frac{d^2\alpha}{dt^2} = (B+C)\frac{d\beta}{dt}. \quad . \quad . \quad (5)$$

If the gyro has a deviation angle α,

$$\frac{d\beta}{dt} = \omega\cos\lambda\sin\alpha \text{ from (2)}. \quad . \quad (6)$$

Substituting in (5), treating α henceforward as a small angle, and writing α for $\sin\alpha$,

$$H\frac{d^2\alpha}{dt^2} + (B+C)\omega\alpha\cos\lambda = 0. \quad . \quad (7)$$

Equation (7) shows that the compass swings in azimuth with an oscillatory motion of period

$$T = 2\pi\sqrt{\frac{H}{(B+C)\omega\cos\lambda}}. \quad . \quad . \quad (8)$$

(ii.) *Damping of the Oscillations.*—No account has yet been taken of the damping arrangement, without which the compass (being only very slightly damped by bearing friction) would oscillate perpetually about the meridian. To prevent this, the point s is set slightly to the east of the vertical line through a. Let r = the angle $f_1 as$, a small

[1] For a general account of the theory of a gyrostat see article "Gyroscope," Vol. I.

[2] See article "Gyroscope," § (6), Vol. I.

angle of about 1°. Then the couple due to the bail may be resolved into components approximately

$B(\beta+\beta_0)$ about the horizontal tilt axis c,
$rB(\beta+\beta_0)$ about the vertical axis ff_1

The latter affects the rate of change of β, so that while equation (5) remains unaltered, (6) becomes

$$\frac{d\beta}{dt}=\omega\cos\lambda\,.\,\alpha-rB\frac{(\beta+\beta_0)}{H},$$

and instead of (7) we have

$$H\frac{d^2\alpha}{dt^2}+r\frac{Bd\alpha}{dt}+(B+C)\omega\cos\lambda\,.\,\alpha=(B+C)r\omega\sin\lambda. \quad (9)$$

Thus in the steady state, when $d^2\alpha/dt^2$ and $d\alpha/dt$ are each zero, the compass has a deviation

$$\alpha_0=r\tan\lambda.$$

This deviation is due to the method of damping, and is therefore known as the "Damping Error" or sometimes as the "Latitude Error," because it varies as the tangent of the latitude angle. It is corrected in the readings shown by the compass, by shifting the lubber's mark out of the fore-and-aft line by an appropriate amount.

Since in equation (9) r^2B^2 is numerically less than $4H\omega\cos\lambda\,(B+C)$, the solution is

$$\alpha-\alpha_0=Ae^{-rBt/2H}\sin\left(\frac{B+C}{H}\omega\cos\lambda-\frac{r^2B^2}{4H^2}\right)^{\frac{1}{2}}t,$$

the equation of a damped harmonic oscillation with period

$$T_d=\frac{4\pi H}{\{4(B+C)H\omega\cos\lambda-r^2B^2\}^{\frac{1}{2}}}.$$

The ratio of successive swings is given by

$$f=e^{\frac{-rB}{4H}T_d}.$$

The value of f is normally about $\frac{1}{2}$ to $\frac{1}{3}$.

(iii.) *Speed Error.*—The above applies to a compass stationary ashore. In a ship moving over the earth's curved surface with a northerly velocity v miles per second at the equator, the apparent axis of the earth's rotation is shifted by an angle δ whose tangent is $v/R\omega$, where R is the earth's radius in miles. If the ship's speed is V on a course K degrees from north and the ship is at latitude λ, we may write (since δ is a small angle)

$$\delta=\frac{V\cos K}{R\omega\cos\lambda}. \quad . \quad . \quad . \quad (10)$$

If V is given in knots and δ is required in degrees,

$$\delta=\frac{1}{5\pi}\frac{V\cos K}{\cos\lambda}=0{\cdot}064\frac{V\cos K}{\cos\lambda}.$$

This "Speed Error" deflects the compass to the west on northerly courses and to the east on southerly courses. Its amount is about 1° for each 10 knots in British latitudes. It will be noticed that this error is entirely independent of the design of the compass. Speed or course error is inherent in any north-seeking gyro compass, although mechanical means may be arranged to correct it. Such an arrangement has been embodied in the Sperry compass, which has a latitude dial and a dial to be set by hand to the ship's speed. An ingenious mechanism then automatically corrects the readings by shifting the lubber's mark by the amount required for the course the ship is steering.

While the ship is actually changing its northerly velocity, the acceleration dv/dt produces a shift of the apparent vertical, as shown, for instance, by a plumb-line, through a small angle whose circular measure is $(dv/dt)/g$, where g is the acceleration of gravity in miles per sec.^{2}. As the case and bail do not alter their tilt relative to the horizon, this is equivalent to a change in the value of β, the tilt relative to the vertical. The pendulous moments of the bail and case then cause the compass to deviate in azimuth at the rate

$$\frac{d\alpha}{dt}=\frac{1}{g}\frac{dv}{dt}\cdot\frac{(B+C)}{H},$$

which must be multiplied by the cosine of the course if the acceleration is not along the meridian. By integrating over the period during which the acceleration persists (which is supposed short)

$$\alpha=\frac{V(B+C)}{gH}\cos K. \quad . \quad . \quad . \quad (11)$$

This is the "*ballistic deflection*," and for a northerly acceleration it is to the westward, and therefore of the same sign as northerly speed error. By making the amount of this deflection equal to the change of speed error the acceleration will move the compass dead-beat to its new resting position without any oscillation. Thus with the co-operation of the correction mechanism referred to above the compass will indicate true north before, during, and at the end of the acceleration. To secure this desirable result, δ in (10) must have the same value as α in (11), then

$$B+C=\frac{gH}{R\omega\cos\lambda}. \quad . \quad . \quad . \quad (12)$$

By substituting this value of B+C in equation (8) the periodic time of the undamped compass must be $T=2\pi\sqrt{R/g}$, or about 85 minutes.

The damping—for $f=\frac{1}{3}$—increases this period to about 90 minutes, and practically all gyro compasses are now constructed to oscillate in this time.

It will be noticed that the value of B+C in (12) is dependent on the latitude. Previous to 1918 compasses were designed to be correct in latitudes of about 40° to 50°, leaving a residual error outside those limits which grew to be rather serious in high latitudes. A modification of the Sperry compass patented by Commander G. B. Harrison and A. L. Rawlings of the Admiralty Compass Department, and

which will be described below, contains a means of altering the value of B to make the ballistic deflection approximately correct in any latitude.

(iv.) *Further Effects of Change of Course.*—Owing to the eccentric connection between the bail and gyro case (s in *Fig.* 5) the couple produced by the acceleration is not confined to the vertical plane. The component in the horizontal plane deflects the gyro axle slightly from the level position, and this causes the compass to wander slightly back from the correct azimuth after it has been placed there by the ballistic deflection. This disturbance has been named "*ballistic tilt*," and the resulting temporary deviation, which attains its maximum value about a quarter of an hour after the acceleration which produced it, is about 2° for 40 knots change of northerly speed. It may be reduced by making the eccentricity less, *i.e.* by reducing the angle r. This of course reduces the damping also. Sperry compasses, in which each swing was one-third the amplitude of the preceding one, have been altered to make this ratio one-half with marked success. The ballistic tilt effect was halved and no noticeable disadvantage was found with the lighter damping.

In compasses damped by other means, such as those of Anschütz and Brown, a precisely similar deviation has been found due to disturbance of the damping oil. Brown has introduced an automatic cut-out to get over this difficulty.

In the case of the Sperry, the deviation is slightly greater if a turn from north to south or south to north is made through west, and rather less if made through east. This is due to the eccentric connection being swung over as the turn is made in spite of the ballistic gyro. A similar effect, but exchanging east for west, has been observed with the Harrison-Rawlings modification.

(v.) *Recent Improvements.*—The sensitiveness of the ballistic gyro to east-west accelerations is occasionally the cause of serious deviations. If the ship rolls badly on a meridional course, the ballistic gyro may swing quite clear of the grooves in the bail and case and put the compass out of action. This was overcome to some extent by adding a weight to the ballistic gyro frame above its point of suspension so as to increase its natural period and make it more stable, but the Admiralty and the Sperry Co. have now adopted the Harrison-Rawlings modification, which does away with the ballistic arrangement entirely.

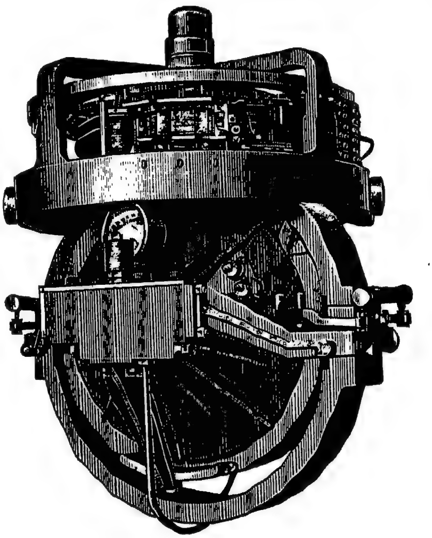


FIG. 6.

In this device the Sperry bail is replaced by a large U-tube containing mercury and opening at the ends into cast-iron boxes (see *Fig.* 6). When this is tilted by the gyro, an excess of mercury accumulates in the lower box and applies a couple which is opposite in sign to that of the pendulous bail. By reversing the spin of the gyro, however, the precession is made in the right direction, and the compass functions as usual. The inertia of the long column of liquid in the U-tube keeps it comparatively steady while the ship rolls and so eliminates rolling error. The above equations for the Sperry compass can be applied to the Harrison-Rawlings modification by making H and B both negative and neglecting C, the pendulousness of the case, which is zero in this arrangement. The quantity denoted by B, which is the "pendulous moment" of this form of bail, depends on the area of the free surface of the mercury to the boxes. A simple device is fitted which makes this area adjustable by turning the knob seen at the right-hand side of the box, and the compass can thus be set to give a correct ballistic deflection in any navigable latitude. With this arrangement the compass has been found proof against any weather, provided that the rolling is not so severe as to swing the gyro against the gimbal rings.

(vi.) *Accuracy.*—In conclusion, some idea may be given of the accuracy to be expected of a gyro compass under average conditions, and the probable causes of any deviations which may be met with. Either the original Sperry compass, or its modification just described, are liable to change of settling-point with a minute change in the balance of the gyro case. An increase of temperature causes the case to expand unequally, making it generally North heavy; change of atmospheric pressure with varying barometric height has a similar effect, as also has any variation of the vacuum maintained in the case. Wear of the bearings on which the wheel spins or on which the case tilts is a further cause of change of balance. The effect of any of these is small, and it would be very unusual for their combined action to change the settling-point more than 2° if the balance was not otherwise disturbed. As the

master-compass is kept below at a fairly constant temperature, no rapid changes of balance are likely to occur. Hence, if no correction by true bearings were available for periods of a week at a time, the compass might be relied on to keep within 1° of its settling position during the interval. Large deviations may be caused by wrong adjustment of the suspension wire by which the vertical ring is supported. This should be installed quite free from twist and checked from time to time to see that it remains so. The deviation due to a twisted suspension is, however, practically constant and does not introduce any uncertainty. The electrical arrangements of the master-compass and the transmission to the repeaters, when entrusted to an intelligent electrician, have proved very reliable, and the gyro compass equipment as an aid to navigation has justified its inclusion in the equipment of modern warships.

§ (16) THE LOG LINE.—In its old form the log line consisted of a wooden board of approximately triangular section, weighted on one side so that it would float upright and present the greatest possible resistance to being dragged through the water. The log line is made fast to the centre of the board and runs out from a reel. The determination of speed is made by noting how many "knots" of the line run out in a definite interval, usually 14 or 28 seconds.

Steamships, in which the course remains constant for great lengths of time, always now use some form of patent log, an instrument which measures distance run rather than speed. A log of this type is made of approximately cylindrical shape with spiral vanes fixed to the curved surface. It is towed astern of the ship, and as it travels through the water the vanes cause it to rotate on its axis. The rotation twists the tow-line and the latter turns a counting mechanism fixed to the taffrail, so that the distance run can be read off. In some patterns the log is practically upon the surface, in others it sinks to some distance below.

If there is a surface current both types show, in general, inaccuracies. A log towing at considerable depth may be completely below the current, and will thus indicate the speed of the ship over the ground; a log towing on the surface would show the speed of the ship relatively to the water, provided the current extended to a depth well below the keel of the ship. In still water both types are equally reliable.

§ (17) METHODS OF POSITION FIXING. (i.) *The Station Pointer.*—One of the commonest methods of fixing position when off a coast is by the measurement by a sextant of the two angles between three recognisable marks on shore.

If A, B, and C (*Fig.* 7) are these points and P the position of the ship, the angles APB (a) and BPC (β) are measured. The position can then be graphically constructed upon the chart by drawing the circle APB to contain the angle a, and the circle CBP to contain the angle β. Since the two circles must have the point B in common, their other point of intersection P is uniquely determined.

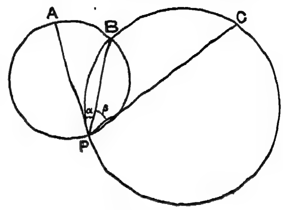


FIG. 7.

The construction of these circles being laborious and difficult when their curvatures are flat, the station pointer is used to achieve the same end. The instrument consists of a graduated circle with three arms, of which the chamfered edges are radii of the circle. One of the arms is fixed and the other two can be rotated, one to the right and the other to the left. The scale of graduations has the zero on the edge of the fixed arm, and the graduations extend from 0° to 180° on either side. The measured angles are set on the instrument by means of the two movable arms, a vernier being available for accurate setting. The instrument is placed upon the chart and moved about until the edges of the legs pass accurately through A, B, and C. The centre of the circle must then be the point P, which can be marked down upon the chart as the ship's position.

Accuracy in the determination of a position by this method is very largely a question of proper choice of the three shore marks.

If it should happen that the angle BCA, as in *Fig.* 8, is equal to the observed angle a, then a circle can be drawn through the *four* points A, B, C, and P and the ship's position becomes completely indeterminate; it may be anywhere on this circle. The two circles of *Fig.* 7 coalesce, and any point on the circumference fulfils the conditions as regards the magnitudes of the angles APB and BPC.

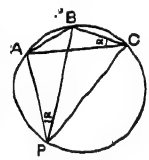


FIG. 8.

The criterion of the suitability of the three marks depends upon the relative positions of A, B, C, and P. In particular, the "fix" is a good one if the middle object is on the nearer side of the line joining the two outside ones. The errors of position obtained by this method are dependent upon the errors of measurement of the two angles, errors of setting the two arms, and errors of locating the station pointer on the chart. Probable values for each of these errors could be obtained and combined

so as to give a probable error of position of each of the two position lines, in this case a small part near P of the circumferences of the circles PBA and PBC. The most probable position is then the point P. The probability of position, per unit area of the chart, diminishes as we move away from P, but diminishes at different rates for different directions. This probability of position per unit area is constant around the circumference of an ellipse whose centre is P and of which the tangent lines at P to the two circles are conjugate diameters (see *Fig.* 9). One of these ellipses is such that the probability of the position lying within it or outside it is one-half. An ellipse of three or four times the linear dimensions gives something closely approximating to certainty of position.

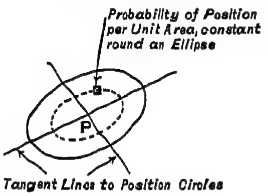


FIG. 9.

It should be noted that similar results hold in all cases where a position is fixed by means of two position lines, each of which has a definite probable error of lateral displacement.

When three or more position lines are found for the purpose of determining a "fix," and they do not pass through a point, the most probable position must be determined by the method of least squares, but it is not possible to make any simple statement as regards the way in which the rate of decrease of the probability of position per unit area varies with variation in direction.

This method of position fixing is almost always used for fixing soundings when making a survey. The sextant used in such case is of special construction, and is called a sounding sextant. (For description see under "Sextants," § (20).)

(ii.) *The Bearing Plate or Pelorus.*—Owing to the impossibility in many cases of obtaining an all-round view from the standard compass, it is necessary at times to employ subsidiary means for taking bearings when the standard compass cannot be used. The bearing plate used for this purpose consists of a circular brass circle, divided into degrees, slung in gimbals so as to remain horizontal when the ship rolls. A concentric circle fits inside the brass one, and is also graduated. This is used like a compass card, and is set so as to show the course by standard compass against the zero of the brass circle, used as a lubber line. The ship's head must of course be kept steady while an observation is being made. A sight vane fits over the brass circle and is used in the same way as the azimuth circle of the standard compass. When a bearing is taken it can be read relatively to the ship's head on the outer circle or relatively to the standard compass on the inner one.

(iii.) *Sounding Machines.*—When soundings are required in water too deep for the ordinary hand lead a sounding machine is used. It consists of a framework supporting a drum round which is wound about 300 fathoms of 7-strand wire. The drum can revolve freely on its axle, can be braked, or can be wound up either by hand or by electric motor. The wire is led from the drum through a block at the end of a spar 30 to 40 feet long, projecting beyond the ship's side, and at the end of the wire is a short length of hemp rope to which a 24-lb. lead is fixed and also a brass guard tube, the latter being about 6 feet above the lead. A dial indicator on the machine shows how much the wire has run out. When sounding from a ship at rest the length of wire run out gives the depth immediately, but when the ship is under way the lead does not sink straight down but trails away aft. In such case the depth is determined by the "chemical tube"—a glass tube open at one end and sealed at the other. The inside of this tube is lined with chloride of silver, coloured red. The tube is placed inside the guard tube open end downwards, and the sea-water has access to it through holes in the guard tube. As the lead sinks water enters the tube compressing the air, and the length of the tube thus filled is connected with the pressure, and therefore with the depth, by a simple extension of Boyle's law. This length is immediately visible when the lead is hauled up, as the part of the chloride of silver that has been in contact with sea-water is turned white and can be measured. As a check upon the accuracy of the sounding, as found by the chemical tube, the amount of wire run out, which is a function of the depth and the speed, can be used.

In very deep water great advances have recently been made by sounding by acoustical methods. If a small charge be exploded near the ship's side the reflected sound wave from the bottom of the sea can be easily detected. With suitable means for accurate measurement of the time between the explosion and the echo it is possible to determine the depth with a very close degree of accuracy.

Another and somewhat similar method is to drop overboard an explosive charge which sinks at a uniform speed which is known. The charge explodes on striking the bottom, and the elapsed time before the explosion is heard leads to a determination of the depth.

§ (18) METHODS OF NAUTICAL ASTRONOMY.—When out of sight of land the navigator has from time to time to obtain his position

by means of observations of the sun or stars. A measurement is made by a sextant of the angular altitude of the heavenly body above the sea horizon, while the time is noted by a chronometer whose error on G.M.T. is known.

Astronomical observations are also necessary to check from time to time the error of the compass. To this end an azimuth mirror is fitted to the standard compass for the purpose of taking the bearing. The time being noted when the bearing is observed, and the latitude and longitude of the ship's position being known with sufficient accuracy, the true bearing of the sun can be calculated and the compass error determined.

Occasionally special instruments are used for the purpose of measuring the depression of the visible horizon below the true horizontal. Tabulated values of this depression are available, but are sometimes inaccurate owing to abnormalities of the refraction.

§ (19) THE SEXTANT. (i.) *Description.*—After the mariner's compass the next most important instrument that the navigator uses is the sextant. With it he measures the angular altitudes of the sun and of the stars above the visible horizon. From these observations his position on the earth's surface is determined. With the sextant also he measures horizontal angles between visible terrestrial objects from which his position when near land can be plotted upon the chart.

Simple in principle and simple in construction as the sextant is, it was not invented until nearly the middle of the eighteenth century, and earlier navigators had to measure altitudes by the cross-staff, a wooden apparatus resembling a T-square with three sliding crosses of unequal lengths. The instrument was held with the long arm of the T horizontal, and the observer looked along this and maintained it so as to be in the direction of the horizon. One of the crosses was then moved until its upper edge was in line with the sun's centre and the altitude read off from a scale of tangents. A modified form was the back staff, in which the observer turned his back upon the sun and held his instrument so that light from the sun passing through one slit fell upon another. If at any time the observer could through this second slit see the sea horizon his sight was correctly made; if not, the slit nearest to the sun needed setting up or down. The cross-staff, or fore staff, and the back staff were the principal measuring instruments in use by all the great navigators of the Middle Ages. Cabot, Columbus, Hawkins, Drake, Raleigh made their great voyages by no other aid than theirs. Chronometers, as we know them, that will keep a fairly steady rate for months on end, had not been designed. The lunar distances of stars were not then tabulated, so that they could make no observations for longitude, and their "sights" consisted merely of the determination of the latitude at noon. About 1664 a reflecting form of backsight was invented by Hooke, but a cursory examination of its design shows that it was not properly a sextant, inasmuch as a small tilt of the instrument in the vertical plane would throw the images of the sun and the horizon out of contact. Presumably for this reason it was never used, and although reinvented by Grandjean in 1732 (*Fig.* 10), and approved

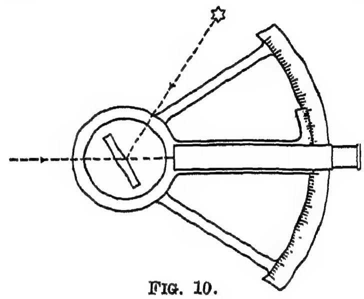


FIG. 10.

by the Académie Royale des Sciences de Paris, it was not until a double-mirror instrument was designed a few years later that the navigator possessed anything corresponding to the sextant as we now know it, an instrument that can be tilted in the vertical plane and yet maintains contact between the image of the sun and the image of the horizon. This condition can only be realised provided both images are formed after an even number, or both after an odd number of reflections. Customarily the sun is reflected twice and the horizon not at all; but several designs have been produced in which each is reflected once, or in which one is reflected once and the other three times. As to who was the first designer of the sextant is not known with certainty. Popularly it is ascribed to Hadley in 1731; but Newton certainly designed one independently about the same time, and there are claims of one designed by an American merchant captain of an even earlier date.

In all these earlier forms of sextant the main principle is shown in *Fig.* 11. It consists of a framework or limb A to which is attached a fixed mirror H, the horizon glass. The limb is in the form of a sector of a circle, and a movable arm R, the index arm, can rotate round the centre of this circle. Fixed to the index arm and rotating with it is the index glass I. The horizon glass is usually silvered over half its area, so that the observer looking horizontally into the horizon glass can see the twice-reflected image of the sun in the silvered portion and then also see the

horizon directly through the unsilvered portion. By using a small telescope T attached to the limb these two images become superposed, and, the telescope being usually of the inverting type, the field of view is as

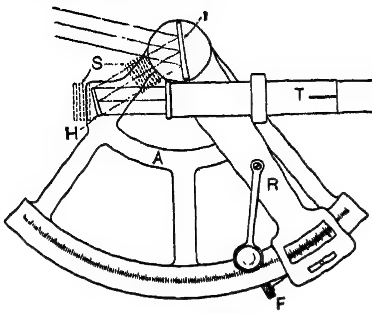


FIG. 11.

shown in *Fig.* 12. When exact contact is made the measure of the sun's altitude is twice the angle between the two mirrors, and can be read off from a scale engraved on the arc of the instrument. By cutting this scale in "sextant degrees"—*i.e.* 720° to the complete circle—the altitude measured can be read direct.

The use of the instrument requires considerable practice before the observer can become proficient. The sun is first of all "brought down" to the horizon. To do this the index arm is set somewhere near zero and the sextant held in a vertical plane, but with the sight line directed towards the sun instead of horizontal. Two images of the sun will then be seen—one in the silvered and one in the unsilvered portion of the horizon glass. The sextant is now brought slowly down to its normal position, moving the index arm all the time while this is being done in order to keep the reflected sun in the horizon glass, until ultimately the horizon appears through the unsilvered portion. "Bringing the sun down" is usually carried out with the telescope removed so as to give the observer a greater angular field of view.

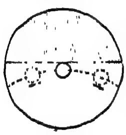


FIG. 12.

When the sun and horizon are both visible in and through the horizon glass the index arm is clamped to the limb and the telescope shipped, the final adjustment necessary to make contact being done by means of the fine-motion screw or "tangent screw" F attached to the arm. The final adjustment in itself requires care. If the sextant is "rocked"—*i.e.* given a small angular rotation round the axis of the telescope—the sun will move relatively to the horizon as shown in *Fig.* 12, and the adjustment is only correctly made provided that in its motion across the field of view the sun swings just to touch the horizon and just does not dip into it. To keep the sun in the field of view while this rocking motion is being made, the whole instrument has to be suitably turned in azimuth. By acting as described above, the navigator makes a measurement of the apparent altitude of the sun's lower limb above the visible horizon with an accuracy of observation normally of about 30″.

The sextant has, in addition to the parts described, other fittings that need to be referred to. Shades S are placed between the two mirrors in order to reduce the brightness of the sun. They consist of pieces of dark glass with parallel faces and of different densities. If not made with their two faces exactly parallel they will refract the light passing through them, and a consequent error in the measurement results. With a shade that is prismatic this deviation will be greatest if the edge of the prism—*i.e.* the line of intersection of the two plane faces—is at right angles to the limb, and will be zero if the edge is parallel to the limb. As the shades need to be mounted loosely in their frames on account of the variation in temperature that a sextant has to withstand, and as the circular form is the most convenient to make, the shades have ordinarily to be made of glass so nearly parallel that light passing through it is not deviated by more than 5″. The shades are usually four in number and of densities such that the transmission ratios are 1 in 40, 1 in 400, 1 in 7000, and 1 in 35,000, but such figures are very approximate.

Three shades and sometimes four are also placed behind the horizon glass. The sextant has at times to be used with a second image of the sun reflected in a mercury trough, instead of with the sea horizon. There is also frequently a good deal of reflected light from the surface of the sea, and a special light shade is included to cut off some of this glare. Sometimes also a Nicol prism is placed behind the horizon glass to cut out the polarised light reflected from the water.

(ii.) *Errors.*—The sextant in use is always subject to errors of varying amounts. The mirrors are not rigidly held, and change their position; and an examination of the value of the zero error, which is the most important of the first-order errors, is necessary from day to day. The determination is made by the sea horizon, or by the measurement of the sun's diameter "on and off the arc." Another first-order error is that of centring, the scale being cut from a point which is not the working centre. Sextants are tested in this respect at the National Physical Laboratory. The amount of decentration which is allowable in order to comply with the N.P.L.

test, that the error must not exceed 40 seconds of reading, lies between about one-thousandth of an inch in the direction of the middle reading and about three-thousandths in the direction at right angles, with a sextant of 7-inch radius. Whether the value of the centring error as given in the N.P.L. certificate is permanent is a very doubtful point. The sextant in use is exposed to hot sunlight on one edge of the limb while the rest is in shadow. Some sort of deformation of the limb must take place, and it is quite possible that an additional minute or two of centring error is introduced every time the sextant is used in the tropics. Makers of sextants have frequently contributed to the risk of this trouble by finishing the limb a dull black, but of late there has been a tendency to make sextants in which the metal work is much less absorbent of heat.

First-order errors also arise from the shades and mirror being made from non-parallel glass. The worst effect on the reading occurs when the "edge" of the prism is perpendicular to the plane of the limb, no error being caused when this edge is parallel to the limb. If shades were mounted so as to be non-rotatable the prismatic error would always be of the same magnitude and might be included in the zero error, although it would involve a separate correction for each shade.

A prismatic horizon mirror gives rise also to a constant error, but with a prismatic index mirror the resulting correction is different for each reading and increases rapidly towards the higher angles. As a consequence it is necessary that mirrors should be made with a high order of accuracy in parallelism. In addition to absolute error there is also trouble in the duplication of an image if the mirror is of non-parallel glass, one image being formed by light reflected from the first surface of the mirror, the other by light reflected at the back. A simple practical test for parallelism of the mirrors is obtained by observing through the highest-power telescope the image of a star seen after reflection at the two mirrors, the index mirror being set to the top end of the arc. If under these conditions no duplication of the image is noticeable the mirrors are quite good enough for the purpose.

The second-order errors of a sextant are due to three principal causes, viz. want of perpendicularity of the two mirrors to the plane of the limb and want of parallelism of the line of sight to the same plane. When all these three defects occur together the resulting error in the reading can be shown to be

$$\begin{aligned}\epsilon = 2\operatorname{cosec} R\{&\alpha^2\cos(d+\tfrac{1}{2}R)\cos(d-\tfrac{1}{2}R)\\ &+\beta^2\cos d\cos(d-\tfrac{1}{2}R)-\gamma^2\sin^2\tfrac{1}{2}R\\ &-2\beta\gamma\sin\tfrac{1}{2}R\sin(d-\tfrac{1}{2}R)+2\gamma\alpha\sin d\sin\tfrac{1}{2}R\\ &-2\alpha\beta\cos d\cos(d-\tfrac{1}{2}R)\},\end{aligned}$$

where R is the value of the reading, d the "sextant angle," viz. the angle between the normal to the horizon mirror and the axis of the telescope, and α, β, γ are the errors of the index mirror, horizon mirror, and line of sight respectively. If R is zero the value of ϵ is not zero unless $\alpha=\beta$. This fact is clearly obvious, as, without this condition, the two mirrors cannot be moved so as to be parallel to one another, and the two images of the same point cannot be made to pass over one another.

All sextants are provided with means to adjust the second-order errors. The procedure is, usually, to set the index mirror so that its plane approximately bisects the arc. One end of the arc can then be seen directly and the other by reflection in the index mirror. The latter is then adjusted until the two views of the arc are coincident. When carefully done the value of α should not exceed two or three minutes. Having set up the index mirror, the horizon mirror is then adjusted so that the two images of a distant point can be made to pass over one another. As the telescope can be used for this purpose, α and β can be made equal to within ten seconds or so.

The adjustment of the collimation error is much more troublesome. The telescope is sometimes mounted in an adjustable collar with two small setting screws, but the adjustment is hard to make at sea and, indeed, is not often attempted. The weakness of the rising piece and the unsubstantial design of the framework of the limb are so considerable in some sextants that slight pressure between the ocular end of the telescope and the side of the nose may cause a temporary collimation error when the instrument is in use. In addition the observer may take his observation away from the centre of the field, so that the value of γ is of necessity higher than α or β. If we take α and β equal and comparatively small relatively to γ, the approximate value of the error is

$$=-\gamma^2\tan\tfrac{1}{2}R+4\gamma\alpha\sin\tfrac{1}{2}R\cos(\alpha-\tfrac{1}{2}R)\sec\tfrac{1}{2}R.$$

The value of this is negligible at small angles. At 90° it becomes $-\gamma^2+2{\cdot}15\gamma\alpha$, and at 130°, which is about the maximum observable angle, $-2{\cdot}14\gamma^2+4{\cdot}85\gamma\alpha$.

It will be seen from the above that if the errors of the mirrors are carefully eliminated the error of the reading is inconsiderable except at the high angles, and that for sea work, where altitudes do not exceed 90°, collimation error of a degree is not serious. On the other hand, if the mirrors are out of adjustment an error of reading of four or five minutes is quite possible.

§ (20) FORMS OF SEXTANTS.—No very great change has been made in the design of sextants since the time of Hadley. With improvements

in the methods of cutting the scale and with the use of more accurate verniers for reading, it has been possible to reduce the radius of the limb without any loss in accuracy. Older patterns frequently had a radius of as much as 15 inches as against the 7-inch radius that is customary nowadays. Telescopes have been improved and have now better definition and larger field of view than formerly, but there is a limit to what is possible in the latter respect owing to the difficulty in increasing the size of the mirrors. Some improvement is possible if the mirrors are replaced by reflecting prisms in which the refractive effect of the glass can be utilised to give an enlarged aperture. For example, if a prism of angles 45°, 45°, and 90° (see *Fig.* 13) be used, a reflection can be obtained when the incident light is parallel to the reflecting face and at the same time the breadth of the reflected beam is about a quarter of the length of the reflecting face. If a mirror had been used in this way the breadth of the reflected beam would be *nil*.

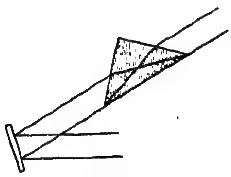


FIG. 13.

A prism used for reflection in place of a mirror need not have its angles exactly 45° and 90°, but the two base angles should be equal. For ordinary sea work the difference between the two base angles should not exceed half a minute, although an error up to ten minutes is permissible provided a suitable correction is applied to the reading. The value of the latter is very similar in its magnitudes to a centring error, and if the centre is made adjustable it is quite possible to throw the working centre out so that the true centring error thus introduced neutralises the prism error and leaves a balance of error below twenty seconds for all points of the arc.

The use of a prism in place of a mirror on the index arm gives another advantage. In order to obtain a wide range of readings (up to 130°) the sextant angle needs to be kept down to about 15°. This involves considerable space being left between the end of the telescope and the horizon mirror and between the latter and the index mirror in order to avoid any cutting off of light. The more these distances are increased the more difficult it is to obtain a wide external field of view. With the use of a prism to replace the index glass the sextant angle can be made larger and the spacing of the different parts much reduced without any loss in the working range of the sextant, and with actual gain in aperture at the high angles. Modern improvements in the production of prisms may possibly result in their replacing mirrors entirely in sextants.

Another direction in which tentative efforts at improvements in design have been made lies in the method of reading the arc. Under good conditions of illumination and with plenty of time to spare there is no great difficulty in reading a 7-inch sextant to ten or twenty seconds, but at night-time at sea it is much more difficult, and efforts have been made on the Continent to provide the navigator with a sextant which is easier to read and on the whole good enough for his needs. The usual form that this improvement has taken is the provision of a worm gearing round the edge of such a pitch as will give 720 teeth to the complete circle. The head of the worm is divided into sixty parts and a reading can be made with great ease. No great difficulty is experienced in obtaining accuracy within a minute, although there is always danger of burring the teeth of the gear by accident. A circular rack-and-pinion movement has also been tried for the same purpose, but not with the same success.

In addition to the ordinary sea uses of measuring the altitudes of heavenly bodies, sextants are sometimes required for the accurate determination of latitude and longitude of places on shore, or for the determination of the errors of chronometers at a place whose position is accurately known. In such cases the measurement of the altitude is made between the sun or star and its reflected image in a mercury trough. Usually a higher degree of accuracy is required with such an observation and the sextant is fitted with a telescope of higher power and has an arc of larger radius cut sometimes to 5 seconds. For the purpose of this type of observation the sextant is usually mounted on a stand, which is a light three-legged pedestal with a universal joint at the top to which the sextant is clamped. After an observation has been made a sextant has to be turned up to take the reading, and various stops are sometimes fitted to the stand so as to permit of the sextant being immediately turned back again to precisely the same position as before. In order to reduce errors of observation the navigator generally likes to make three observations at sea and at least five on shore, the mean of the altitude being assumed to correspond with the mean of the time. To avoid delay in having to read the sextant after each observation, various devices have been employed. In one such device the tangent screw is fitted with a ratchet and the pitch of the screw chosen so that each tooth of the ratchet is equivalent to one or two minutes change in altitude. With such an arrangement the navigator can obtain a set of observations

and does not need to read the sextant until they are completed.

Another device is by the use of a reflecting prism of small angle which is placed beyond the horizon glass. The edge of this prism is perpendicular to the plane of the limb, and according to whether the prism is placed base up or base down the horizon is lowered or raised. A set of three observations can be made (i.) by using the prism base up (see *Fig.* 14), (ii.) by removing it entirely (see *Fig.* 15), and (iii.) by using it base down (see *Fig.* 16). The mean of the three altitudes will be the sextant reading irrespective of the angle of the prism.

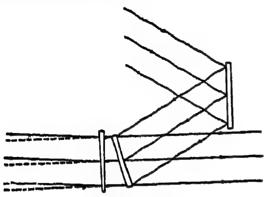


FIG. 14.

Another use to which sextants are frequently put is the measurement of horizontal angles between two fixed marks. In a hydrographical survey the positions of soundings are usually fixed in this way, two horizontal angles between three fixed objects being measured as nearly as possible simultaneously. The position of the sounding is subsequently plotted on the chart by means of a station pointer set at the measured angles. The sextant used for this purpose differs only slightly from the ordinary pattern. It usually has a shorter radius as a smaller degree of accuracy is necessary, the shades are abolished, and the telescope is of the Galilean variety so as to give an erect image. Efforts have been made to give the observer as wide a field of view as possible and also to make the sextant easy to read. Various types of "double angle" sextants have been tried in which the two angles can be measured at the same time, so as to avoid the objection of a lapse of time while the first angle is being read.

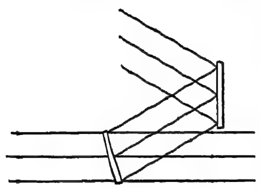


FIG. 15.

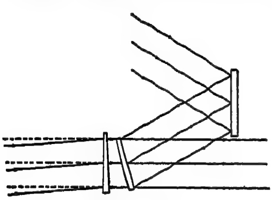


FIG. 16.

§ (21) BUBBLE AND PENDULUM SEXTANTS.—Owing to the difficulty that is frequently experienced at sea of seeing any horizon, attempts have been made many times to replace the sea horizon by some form of level or vertical that is controlled by gravity. In some of these devices the method adopted is to reflect into the field of view the image of the bubble of a spirit-level which indicates the direction of the vertical (see *Fig.* 17). The observation is made by bringing the centre of the sun to the centre of the bubble. In others a horizontal line moves up and down in the field of view under the control of a pendulum, and is exactly opposite a fixed pointer when the telescope is pointing towards the true horizon, so that when the sextant is held true and steady, observations of the sun can be made down to this artificial horizon in the same way as it is usually brought down to the visible sea horizon.

Of the two arrangements the bubble is preferable as it is self-damping if the liquid is sufficiently viscous, and there is an additional advantage that, with proper optical arrangements, the images of the sun and bubble will follow one another across the field even when the instrument is tilted in a vertical plane. Such an instrument can properly be called a sextant. The pendulum instruments are, more strictly speaking, clinometers.

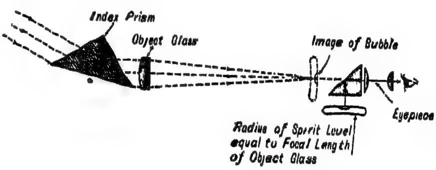


FIG. 17.

All sextants of this type are very severely handicapped by the movement of the ship and consequent horizontal acceleration of the observer and the sextant together. The bubble or pendulum sets itself under such circumstances not in accordance with the true vertical but with the false dynamical vertical compounded of earth's gravity and horizontal acceleration. Horizontal accelerations of 1 ft. per sec. per sec. are common. Such accelerations are only very approximately periodic, and the observer finds that if he brings sun and bubble together at noon, the former will wander up and down relatively to the latter by the amount of a degree or two on either side, and generally so slowly that it is impossible to form even an approximate idea of the mean position. Any method of damping that is practically available serves only to damp out the rapid vibrations of the observer's hand and leaves the slow oscillations untouched. In ordinary weather in the Atlantic, where the ship necessarily rolls and pitches,

and has further alternating horizontal accelerations due to the "send" of each succeeding wave, the probable error (even chance error) of a single such observation taken from the bridge may be put at about 20 minutes of arc. This must be multiplied by four to give the navigator something approximating to certainty, with the result that, having taken such an observation and laid down his position line on the chart, he has then to draw two others eighty miles on either side, and his "sight" tells him that it is reasonably certain that the ship's position is somewhere on the 160-mile strip so drawn.

The above statement explains how it is that bubble sextants which have an accuracy of two or three minutes when used on shore, and perhaps ten or fifteen minutes when used on a cross-channel voyage where alternation of acceleration is more rapid, fail entirely when used in open sea. Nothing can possibly succeed there except a gravity controlled pendulum whose period of oscillation is considerably longer than the periodicity of the ship's acceleration, and the only practical form is some sort of gyrostat.

One such device has already been tried with a certain amount of success, Admiral Fleuriais' gyro sextant. The mechanism is complicated and need not be described in detail. The gyro is spun by an air blast from a pressure tank pumped up by hand. An optical arrangement produces in the field of view of the telescope a spot of light which is reflected from the case of the gyro. The speed and weight of the gyro are insufficiently great to give the requisite length of period, and although the probable error of a single observation is distinctly lower than 20-minute value of the bubble, it is still not low enough for the instrument to be of real practical utility.

There appears to be no reason why the master gyro compass should not be modified so as to give an approximately horizontal level that can be reproduced on the bridge by the repeater mechanism. A mirror attached to the repeater could then be used in a similar manner to the mercury trough used on shore.

§ (22) The Artificial Horizon.—When taking sights on shore with a sextant the angular measurement is made between the sun or star and its image in a trough of mercury. The latter has usually a surface of about 3 by 5 in., and is covered by a glass shade to protect it from being disturbed by the wind. The glass of the shade should be made with its faces accurately parallel, but owing to the difficulty of doing this it is customary to eliminate the error by taking two sets of observations, the shade being turned end for end between the two sets.

Disturbance of the mercury surface can be prevented by using a very shallow pool of it upon a flat copper plate which has been amalgamed and levelled. Surface tension in such case permits of the use of a pool less than a millimetre deep, and if the copper plate is amalgamed only on its upper surface there is no risk of the mercury running off.

This arrangement is also very suitable for cleaning the mercury, which can be done by passing a glass rod lightly over the surface, when all the impurities of the mercury are collected on the rod.

§ (23) The Azimuth Mirror.—In order that the navigator may be able to take bearings of distant objects on the horizon or of heavenly bodies, the standard compass is usually fitted with a special arrangement to enable him to do so. The earliest form of this consisted merely of a foresight and backsight that could be trained on the object. The reading of this direction on the compass card could be approximately accomplished at the same moment.

For bearings of the sun a vertical shadow pin was fitted over the centre of the card and the bearing of the shadow read directly.

An improvement of these two methods is the Kelvin azimuth circle, in which a reflecting surface, usually in the form of a prism, is mounted above the compass. This prism is capable of a rotating movement round a horizontal axis, which has to be turned in azimuth so that it is at right angles to the vertical plane through the centre of the compass and the object (terrestrial or celestial) that is being observed. By turning this prism round its axis to a suitable position light from the object is reflected to a direction making about sixty degrees upwards from the horizontal, and the image so seen can be made to appear to the observer to be close to the nearest edge of the prism. In *Fig.* 18 a

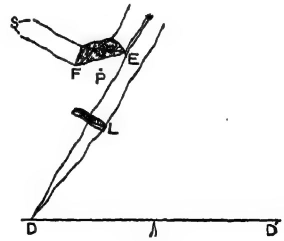


Fig. 18.

vertical section of the arrangement is shown, the plane of the paper being taken to be the vertical plane in which the object lies. The horizontal axis of rotation P is perpendicular to the plane of the paper. Light from the

body S is reflected in the face FE, and the observer can move his head into such a position that the object is seen close to the edge E. Below the prism is a lens L, whose axis also in the plane of the paper passes through the divisions D of the compass card DD′. The focal length of the lens is LD, so that the image of the divisions appears to the observer to be at infinity. Provided the focal length is also equal to the radius of the card, the angular breadth of a division as seen by him will be 1°. This is of importance, as otherwise two objects on the horizon 1° apart cannot be placed against two consecutive divisions of the card at the same time, and a sideways movement of the observer's eye will have a parallactic influence on the bearing as he reads it. Further, it is clear that for observations of objects on the horizon it is not essential that the object shall lie exactly in the plane through the centre of the compass at right angles to the axis of the mirror. For objects not on the horizon, however, the case is different. Two objects, each at altitude α and differing in bearing by 1°, are distant 1° cos α apart, and therefore cover this area on the image of the compass card. When one object appears on a division of the card the other will be at a distance 1° (1 - cos α) from the next division. In theory, therefore, it is necessary to turn the azimuth mirror round so that the object observed lies in the correct vertical plane. If the azimuth mirror is directed K° wrong in azimuth the error in reading will be K° (1 - cos α). The limitations of size of the mirror and lens make it impossible to see the object if K° exceeds 10°, and navigators, as a rule, are chary of trusting to bearings of heavenly bodies of greater altitudes than 30°, but, even so, there is a possibility of error of nearly $1\frac{1}{2}$° in the bearing. With an altitude of 45° this error is increased to 3°.

A further source of error at sea lies in the fact that when the ship is rolling or pitching neither the compass card nor the compass bowl, to which the mirror is fixed, is correctly horizontal, but is probably perpendicular to the false vertical of the moment. Azimuths are then being measured with reference not to the true vertical or horizontal but to the false, and the error made will be zero only when the false vertical is in the plane containing the object and the true vertical.

A second method of using the Kelvin mirror is available for objects on the horizon. This method involves the reflection of the card by the mirror, so that the observer has to look directly at the distant object and read its bearing against the divisions of the card as seen after transmission through the lens and reflection at the prism (*Fig.* 19).

A somewhat similar instrument is the Admiralty azimuth mirror used on the gyro-compass repeater. A vertical sectional view

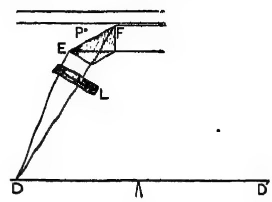


FIG. 19.

of this arrangement is shown in *Fig.* 20. The card is viewed through a lenticular prism A, two faces of which are plane while the third has a curve worked upon it. This prism becomes in effect a lens whose focal length is approximately the radius of the card, so that divisions of the card are projected to infinity

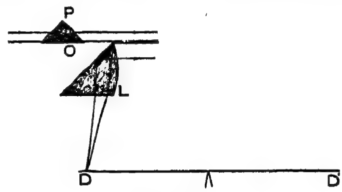


FIG. 20.

and two adjacent marks subtend 1° at the eye. Light from the sun or a star is reflected at the face EF of a prism capable of rotation about an axis perpendicular to the plane of the paper.

As in the Kelvin azimuth circle, errors arise in the case of bearings of objects not on the horizon, the amount of the error being exactly the same as before.

Various other types, similar in principle to the above, are also in use at sea. In all of them, in addition to the errors mentioned, there are defects in construction to be reckoned with which make the total error somewhat more complex than the expression K(1 - cos α) that has been considered. The full investigation of these errors is too long for it to be possible to make a detailed examination of them here.

The azimuth mirror is in constant use at sea for the purpose of obtaining the deviation of the compass when out of sight of land. A "time azimuth" observation is made, and the navigator calculates the true bearing that the sun or star had at the moment of his sight. To do this with absolute accuracy requires exact knowledge of his position on the surface of the globe, and this exact knowledge he seldom has; but, in general, his position is known to within a few miles, and the true

bearing can be calculated with a precision of a quarter of a degree. If this type of sight could always be made on Polaris or some star very close to the pole, whose bearing changed only slightly during twenty-four hours, very large errors could be allowed in his knowledge of his position without appreciably affecting the result of his calculation.

The true bearing being calculated, the observed bearing gives him the total compass error.

§ (24) Measurement of the Dip of the Sea Horizon.—All methods for the measurement of the depression of the visible horizon are based on the assumption that the amount is the same on all azimuths. Hence, if the angular distance is measured between two points on the sea horizon differing by 180° in azimuth, this angular distance will be in defect of 180° by twice the dip.

In *Fig.* 21 is shown the general optical plan of an instrument used specially for the

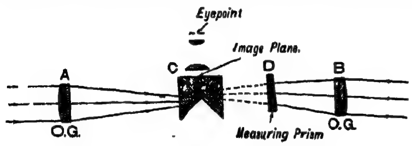


Fig. 21.

purpose. Two object glasses of equal focal lengths are placed at the right- and left-hand ends of the instrument. In the centre are two prisms C, cemented together, which reflect the light coming from the two object glasses upwards. The two object glasses have a common focal plane on the upper horizontal face of these two prisms, so that the cemented interface is seen as a dividing line of the field in sharp focus. A refracting prism P can be moved longitudinally between the right-hand OG and the central prisms, and its motion causes the images of the sea horizon to be displaced relatively to one another (see *Fig.* 22). The prism P can thus be adjusted so as to bring the two sea horizons into line, and the distance of P in front of the common focal plane is thus a measure of the dip. The instrument has to be made slightly more complicated, as with this arrangement it is not possible to measure down to zero depressions. A modification is made by interposing a fixed refracting prism between A and C, and by this means bringing the zero position of P to a point between C and B.

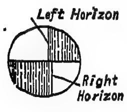


Fig. 22.

Another method of measurement of the dip is by means of a prism, designed by Lieutenant Blish of the U.S. Navy, which can be attached to the sextant, as shown in *Fig.* 23. If the two reflecting faces, A and B, are exactly at right angles to one another the image of the back horizon, as seen by way of the prism,

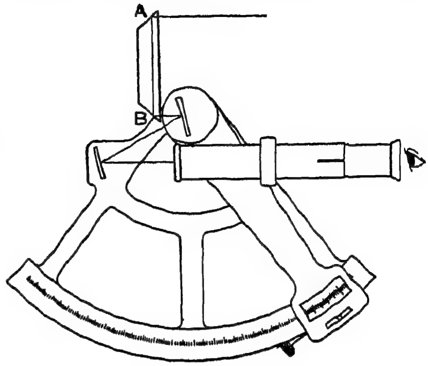


Fig. 23.

over the observer's head, is displaced 180°. The observer consequently sees two images of the sea horizon—one erect and the other inverted—which he can bring together by adjusting the index arm of the sextant. If there is no zero error of the sextant the sextant reading under such circumstances is twice the dip. In actual use this reading has to be corrected for zero error and for error of the prism from 90°.

§ (25) Methods of Directional Wireless Telegraphy.—During the European War the development of the detection of the direction of travel of wireless waves made very great advances.[1] Directional wireless stations were set up in various parts of the country, and by their means the direction in which a wave reached a station could be determined with an accuracy of the order of half a degree, even when the intensity of the wave was only very small. When two stations, say a hundred miles apart, picked up, directionally, the same wireless signal, a short comparison between the two observed bearings enabled the position of the sending station to be fixed with an accuracy of a mile or two, and a vessel at sea in doubts as to her position had only to send out a wireless signal for the shore stations to pick up and to communicate the position back to her.

In the earlier part of the war this method was used by Germany for the purpose of navigating her Zeppelins across the North Sea during their raids on England and Scotland; but ultimately the method was abandoned, as British stations, also picking up the signals, had a more accurate knowledge than Germany of an airship's position. British warships in the North Sea were in a similar situation in

[1] See "Wireless Telegraphy Apparatus," § (11), Vol. II.

respect of the German directional stations, and it soon became highly dangerous for a British ship to use wireless except in close proximity to our own coasts. Within fifty miles or so of our own stations, however, directional wireless had a very extensive use even in wartime, and ships were constantly given their position and guided into harbour in thick weather.

Owing to the danger, from a military standpoint, of the ship at sea sending out wireless signals, attempts were made to install a directional receiving apparatus on board, but not with complete success. The passage of the wave across the metallic structure of the ship produces a local disturbance in direction, so that the "wireless compass" has to reckon with a table of deviations more complex than those of the magnetic compass. In addition the effects of the electrical installations on board, being at very short range, may completely mask the reception of signals.

With aircraft greater success was obtained in the use of the directional wireless receiver, and towards the end of the war the larger machines were equipped with navigational apparatus that enabled them to fix their positions by "cross-bearings" of any stations that they were able to recognise.

Several forms of receiving apparatus were tried. In one of the earliest the receiver was merely a circular or rectangular coil in a vertical plane capable of orientation in azimuth. The maximum current is set up in such a coil when its plane lies in the direction in which the wave is travelling. By turning the coil until the noise in the telephone receiver connected with the coil has its maximum intensity the direction of the signal is determined. It is impossible, however, to do this with anything like sufficient accuracy. No difference in intensity can be noticed for quite 20° on either side of the maximum position.

When the plane of the coil is at right angles to the direction of travel of the wave the intensity is zero, but it is not possible to pick up this direction by trying for silence in the telephone receiver, as on account of the noise of the engines nothing at all can be heard over a considerable angular range.

Two coils were then tried, with planes both vertical and at right angles to one another, which formed a primary and a secondary circuit. Suitable switches made it possible to listen-in on one coil or on both. The primary was first directed as nearly as possible so as to lie in the plane of travel of the wave by listening for maximum intensity. If, in such a position, the current in the primary is A, and that in the secondary B, then by a suitable arrangement of switches the current A + B or A − B can be passed through the telephone receiver, and the ear has a much nicer discrimination of the difference in the two intensities than in the search for a maximum. It was found possible to pick up in this manner the direction with a precision of about half a degree under the best conditions.

In determining the actual direction in which the wave is received a magnetic compass is also needed. The bearing of the wave relative to the fore-and-aft line of the machine is read on a circular scale attached to the movable coils. This bearing requires correction for wireless deviation, which is approximately of the quadrantal type. The compass direction of the machine's head requires the ordinary corrections for magnetic deviation and variation. When flying directly towards or away from a station the magnetic compass is unnecessary, as the machine has only to be steered so that the main coil is in the fore-and-aft line when correctly receiving the bearing. This statement requires modification if the machine is making leeway. In such case, although the machine's head is always directed towards the station, her actual track will be of a spiral form, and she will ultimately reach the station coming up wind.

§ (26) Position Fixing by Directional Wireless.—When making a fix by cross-bearings of two visible objects, all that the navigator has to do is to draw through them on his chart straight lines in the requisite direction, and his position is their point of intersection. Visible distances do not exceed twenty miles as a general rule, and for that distance the effect of the earth's curvature is inappreciable. For wireless cross-bearing distances may go up to a thousand miles, and for such ranges the curvature effect is considerable.

Different means of taking account of this curvature effect have to be adopted according as whether the directional receiving apparatus is in two fixed shore stations or in the ship or aeroplane. In the former case the simple method is to plot the bearings on a gnomonic chart upon which great circles are represented by straight lines. A special "rose" has to be constructed round each station, as in such a chart, except at the centre, angles are not the same as those measured on the earth's surface. As wireless waves travel, in general, on great circles upon the earth's surface, the two straight lines drawn through the two stations to correspond with the two received bearings intersect in the position of the ship.

In the case of directional reception by the ship the problem is more complicated. Let B (see *Fig.* 24) be the position of the sending station and C that of the ship. Ordinarily the latter will be using charts constructed on a Mercator projection so that the great-circle track of the wireless wave will be BGC, a

curve which is hollow on the side towards the equator.

The straight line BRC is the rhumb-line track from B to C.

It will be seen from the figure that the direction at which the wireless wave starts, viz. TBN with the meridian, differs from the direction at which it arrives at C (T′CS with the meridian), which is the direction that C receives, this difference being the sum of the two angles TBR and T′CR.

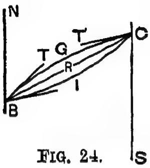


FIG. 24.

For ranges not exceeding 500 miles the Mercator projection of a great circle may be considered to be of constant curvature along its length, so that the two angles TBR and T′CR are equal and of a value which may be shown to be approximately ½ diff. long. × sin mid-lat.

The ship's navigator usually knows his position within thirty or forty miles, so that the difference of longitude between B and C is known to less than a degree. Hence it is possible to estimate to within a degree the value of the "conversion angle" T′CB that will alter the great-circle bearing that he receives to the rhumb-line bearing that he can plot. In this manner he is able to fix his position upon the Mercator chart.

Strictly speaking, having constructed the rhumb line, the navigator is not entitled to assume that his position is somewhere upon it. The *locus* of all points on the earth's surface for which B has the same true bearing is a curve whose shape is more or less that shown by BIC. Within the ranges under consideration this isoazimuthal (or isaz) curve has the same curvature as the great circle, but moves in the opposite sense, so that BGC and BIC are reflections of one another in the rhumb line.

The navigator's position line is consequently not a piece of the great circle CG, nor of the rhumb line CR, but of the isaz CI. For all practical purposes, and within the limits of accuracy with which bearings can be taken, the rhumb line is good enough. For distances over 500 miles and up to 1000 miles the curvatures of the great circle and the isaz curves can no longer be looked upon as constant along their length, and a second-order correction is necessary dependent upon the square of the range.

In addition to errors in directional wireless bearings due to lack of sensitivity of the receiving apparatus, to quadrantal deviation due to the structure of the ship or aeroplane, errors arise from a species of refraction effect when the track of the wave runs more or less along a coast line. No very precise statement can be made as to the amount of this effect, but as a general rule there appears to be a drag on that part of the wave which passes over the land as compared with that part which passes over the water. It would be anticipated, for instance, that a wireless bearing of Poldhu in Cornwall, taken from a position near Dover, would come in from a direction more south than it ought to be. At the same time it is difficult to dogmatise on the point, as possibly intervening land, like the Isle of Wight, might alter the effect completely.

This refractive effect has often been noted, and has usually been found to be more marked at ground level than up in the air. On Salisbury Plain, for example, bearings of Stonehaven taken from the ground are ordinarily about 2° in error. From an aeroplane at 2000 feet the error has almost disappeared.

In addition to errors resulting from the configuration of the land, serious atmospheric effects have to be reckoned with. Their causes are extremely obscure, and a great deal of observation and investigation is still necessary before the navigator can make any allowance for them. At times there appear to be definite tracks of the atmosphere, sometimes in the form of long lanes extending for considerable distances, along which wireless waves travel with greater ease than in any other direction. At times of sunset and sunrise very serious deviations occur, errors of 20° in direction being of common occurrence.

Such disturbances are far more serious than lack of sensitivity of the receiving apparatus, and probably it will not be for some years that the causes will have been sufficiently investigated for the method of directional wireless bearings to be completely sufficient for a ship on a long ocean voyage. Apart from this question, and apart from the quadrantal deviation produced by the metallic structure and the unscreened electrical gear on board, it would seem that the sensitivity of the receiving apparatus has already advanced to the point that the probable error of a single observation is of the order of half a degree, and that a ship fixing her position by cross-bearings of three stations each 500 miles distant can do so with a certainty of being upon a circle of some four or five miles radius.

T. Y. B.

BIBLIOGRAPHY

Astronomical Methods

Admiralty Manual of Navigation.
American Practical Navigator: Bowditch.
Cours de navigation: E. du Bois.
Description et usage des instruments nautiques: M. E. Guyou.
Étude sur les courbes de hauteur: G. Hilleret.
Glossary of Navigation: Rev. J. B. Harbord, R.N.
Handbuch der nautischen Instrumenten (*Hydrographisches Amt der Admiralität*).
Lehrbuch der Navigation (*Reichs-Marine Amt*).

Lehrbuch der Navigation: M. F. Albrecht and C. S. Vierow.
Modern Navigation: Rev. Wm. Hall, R.N.
Navigation and Nautical Astronomy: Staff-Commander W. R. Martin, R.N.
Navigation and Nautical Astronomy: Rev. F. C. Stebbing, R.N.
Practice of Navigation: Lieut. H. Raper, R.N.
Spherical and Practical Astronomy: Wm. Chauvenet.
Text-book on Navigation and Nautical Astronomy: J. Gill.
Treatise on Navigation: J. Merrifield.
Wrinkles in Practical Navigation: Capt. S. T. S. Lecky, R.N.R.

The Magnetic Compass

Admiralty Manual of Deviations of the Compass.
Compensation du compas: L. Joubert.
Der Kompas an Bord (*Deutsche Seewarte*).
Deviation of the Compass: A. W. Reinold and J. W. Waghorn.
Deviation and the Deviascope: C. H. Brown.
Deviations of the Compass in Iron Ships: Capt. E. W. Creak, R.N.
Lectures on Compass Adjustment: Capt. W. R. Martin, R.N.
Magnetism and Deviation of the Compass in Iron Ships: T. L. Ainslie.
Magnetismus und Deviation der Kompasse in eisernen Schiffen: H. A. Jungclaus.
Mariner's Compass in an Iron Ship: J. W. Dixon.
Navigation and Compass Deviation: Commander W. C. P. Muir, U.S.N.
Pivots and Caps in Compasses: R. J. Bray.
Reports of the Liverpool Compass Committee, 1856–57–1861.
The Standard Compass: J. Daniel.
Traité des déviations du compas: A. Madamet.
Traité théorique et pratique de la régulation du compas: A. Collet. (Translation of above by W. Bottomley.)
Various Papers in *Trans. Roy. Soc.*, 1860-61: Smith and Evans.

The Gyro Compass

Foucault, *Comptes Rendus*, Sept. 1852.
Lord Kelvin, *British Association Report*, Montreal, 1884.
H. Crabtree, *Spinning Tops and Gyroscopic Motion* (Longmans, Green & Co.), 1914.
Sir G. Greenhill, *Report on Gyroscopic Theory* (No. 146 of Reports and Memoranda published by the Advisory Committee for Aeronautics) (Wyman & Sons), 1914.
Klein and Sommerfeld, *Theorie des Kreisels.*
Ballistic Deflection of the Gyro Compass, Pamphlet issued by the Admiralty Compass Observatory, Slough, 1917.
Chalmers, *The Gyro Compass*, 1920.

The Sextant

Manual of the Sextant: O. W. Thompson.
Sextant Errors: Naval Instructor T. Y. Baker, R.N.
The Book of the Sextant: A. J. Hughes.

General

Admiralty Manual on the Station Pointer.
Notes on Deep Sea Sounding: Staff-Commander J. E. Davis, R.N.
Sounding Voyage of H.M.S. Hydra: Capt. P. F. Shortland, R.N.

Newton's Rings: the circular interference fringes formed between two surfaces of different curvature when illuminated. They were described by Newton, *Optics*, 1704, and explained on the undulatory theory by Thomas Young, 1802. When formed by monochromatic light they can be employed to determine the curvatures of the surfaces. See "Spherometry," § (7); "Light, Interference of," § (8); "Wave-length, Measurement of," § (1).

Nicol Prism. See "Polarised Light and its Applications," § (10).

Nodal Points and Planes of a Lens. See "Objectives, Testing of Compound," § (1); also "Lenses, Theory of Simple," § (7).

— O —

Objective, Kinematograph, construction of, and aberrations requiring reduction. See "Kinematograph," § (9).

Microscope. See "Microscope, Optics of," § (1).

Telescope. See "Telescope and Telescope Lenses," § (5).

OBJECTIVES, THE TESTING OF COMPOUND

The compound optical systems known as "objectives" may be divided into three main classes, namely, telescope objectives, photographic objectives (or camera lenses), and microscope objectives. The methods of testing their optical constants and general optical performance vary considerably in these three classes, but such variations are to a large extent due to differences in the dimensions (aperture, focal length, etc.) of the objectives. Although the present article is mainly concerned with the first class, namely, telescope objectives, a number of the methods of test to be described will be found applicable to one or both of the other two classes.[1]

The optical performance of any given compound system may be computed by means of theoretical methods if all the optical data of the system are known to a high degree of accuracy.[2]

It is not, however, usually convenient to separate an objective into its components and measure their constants. In the case of an objective in which one or more of the surfaces are locally figured it would be practically impossible to determine its general performance by dissecting it, because it would be difficult, for example, to measure the variations in

[1] For more detailed treatments of these classes see articles on "Camera Lenses, The Testing of," and "Microscope, Theory of the."

[2] For methods of measuring the optical constants of the components of a compound system, such as curvatures and refractive indices, see articles on "Spherometry," "Spectroscopes and Refractometers," and "Immersion Refractometry."

curvature of its surfaces, and even if these could be obtained, the theoretical computation would be very complicated. As a general rule, therefore, it is necessary to employ methods whereby an objective may be tested as a unit.

The chief optical data to be determined in the case of telescope objectives are (1) focal length, (2) aberrations, and (3) general definition.

§ (1) FOCAL LENGTH MEASUREMENT THEORY.—A great number of different methods of measuring the focal lengths of compound lenses have been devised; some of these can only be used in particular cases, while others are of more general application. All the methods are based on the properties of the focal planes, unit planes, nodal points, and conjugate points of an optical system. The fundamental definitions and relations,[1] proofs of which may be found in any text-book on Geometrical Optics, may be summarised as follows. A brief account of them is given in the article on "Lenses, Theory of Simple."

(a) A lens is a portion of a refracting medium bounded usually by parts of two spherical surfaces. The axis of the lens is the line joining the centre of these surfaces.

(b) *Principal Foci.*—There are two points on the axis of a lens known as principal foci, the one in the image and the other in the object space, which possess the properties that

(i.) Rays of light incident on the lens parallel to its axis pass after refraction through a principal focus;

(ii.) Rays of light diverging from a principal focus in the case of a convex lens, or converging to it in the case of a concave lens, are refracted so as to become parallel to the axis.

(c) *Conjugate Points and Planes.*—Conjugate points are two points on the axis such that the image of a small object at the one is formed at the other.

Conjugate planes are planes at right angles to the axis through the conjugate points.

(d) *Unit Points and Planes.*—The unit points are two conjugate points for which the magnification produced by the lens is unity.

The unit planes are planes at right angles to the axis through the unit points.

Thus the image of any point in one unit plane lies in the second unit plane, and the line joining the point and its image is parallel to the axis.

Unit planes are thus planes of unit magnification.

(e) *Nodal Points and Planes.*—The nodal points are two points on the axis such that a ray through the one nodal point, after refraction through the lens, passes through the second nodal point, and is parallel in direction to the incident ray.

The nodal planes are planes at right angles to the axis through the nodal points.

(f) *Principal Points.*—If the media on the two sides of the lens, in the object and image space respectively, be the same, the Unit and Nodal Points coincide, and are known as the Principal Points. (They were introduced by Gauss, who was the first to place on a firm foundation the consideration of lenses and consequently of lens systems.)

The Unit and Nodal Planes similarly become Principal Planes.

(g) The first and second focal lengths, usually denoted by f and f', are the distances between the focal planes and the unit planes in the object and image spaces respectively. If the first and final media are the same, $f=f'$. The cases in which the first and final media are air, are most commonly met with, a notable exception being that of microscope immersion objectives.

(h) If the sizes of an object (normal to the optical axis) and its image are y, y', and the angles which the object and image subtend at the corresponding foci are θ, θ' respectively, then

$$f=\frac{y'}{\tan\theta} \text{ and } f'=\frac{y}{\tan\theta'}. \quad . \quad . \quad (1)$$

(i) If the distances of an object and its image from the corresponding foci are x, x' respectively, and the magnification y'/y be denoted by m, then

$$x=\frac{1}{m}f \text{ and } x'=-mf', \quad . \quad . \quad (2)$$

where the sign notation is the same as is used in the usual system of co-ordinates.[2] From (2) it follows that

$$xx'=-ff',$$

or, in systems where the first and final media are the same,

$$xx'=-f^2. \quad . \quad . \quad . \quad (3)$$

In applying these relations to the practical determination of focal length it is of the utmost importance to be able to find the positions[3] of image planes as accurately as possible. For very rough measurements it is often sufficient to form the image on a piece of ground glass and to estimate its correct position by means of the naked eye. For most purposes, however, it is essential to adopt more refined methods. It is usual to employ a microscope of convenient power for focussing on an image; the power which one chooses depends, of course, on the nature of the object used, the kind of optical system under test, and the accuracy with which the position of the image plane requires to be determined. In the case of certain optical systems there

[1] See also "Optical Calculations,"

[2] In the general case the two focal lengths are assumed to be of the same sign.

[3] For details as to the methods of making some of these measurements, see "Lenses, Testing of Simple."

is a considerable depth of focus and it is sometimes very difficult to focus accurately on the image. In order to eliminate errors due to the range of accommodation of the eye it is advisable, when using a microscope, to employ a Ramsden eye-piece and to place cross-wires or a graduated scale at its focal plane. In the case of a low-power eye-piece, or in making naked-eye observations, the position of focus may be determined by moving a system of cross-wires until no relative displacement between the image and the cross-wires can be detected on moving one's eye from side to side across the field. The position of focus may also be found by covering the aperture of the system under test in such a way that light can only pass through two portions[1] at opposite ends of a diameter. Then two images will be observed unless the microscope is correctly focussed on the image plane. In this case it is assumed that the system is free from spherical aberration. A somewhat similar method, due to Hartmann,[2] is to allow two narrow excentric pencils of light to pass through the system and to measure the distances d_1, d_2 between the centres of the diffusion circles in two planes at some distance on either side of the image plane. If the distances D_1, D_2 of these planes from some arbitrary point are also measured, the distance D of the image plane from that point is given by the relation

$$\frac{d_1}{d_2}=\frac{D-D_1}{D_2-D} \text{ or } \frac{D-D_1}{D_2-D_1}=\frac{d_1}{d_1+d_2},$$

that is,

$$D=D_1+\frac{d_1}{d_1+d_2}(D_2-D_1). \quad . \quad . \quad (4)$$

Another accurate method of determining the position of the image plane of a system was devised by Foucault;[3] in this case an observing microscope is not used. A small illuminated pin-hole is employed as the object and the eye is placed in approximately the position of the image. When a suitable screen, for example a diaphragm with a sharp knife-edge, is moved in front of the eye in a direction normal to the optical axis, the aperture of the system will darken suddenly and uniformly *only* if the screen is in the plane of the image.

In making focal length measurements it is necessary to choose convenient objects; in the case where magnification methods are used the finely divided scales, engraved on glass discs, which are sold by microscope makers are found to be extremely useful. In order to determine the magnification of the observing microscope it is essential to use one of these scales in the focal plane of the eye-piece. For methods which depend on focussing, the most convenient objects are plates of silvered or platinised glass having fine scratches or rulings on the metallic deposit.[4]

§ (2) Focal Length Measurement Experiments. — The chief methods of focal length determination may now be considered in detail. In all cases it will be assumed that the first and final media are air, so that the first and second focal lengths are equal.

(i.) *Methods depending on the Use of Parallel Light.*—(*a*) The position of the second focal plane of an optical system may be found by using an object which is so far away that the rays which come from it may be considered parallel. The sun or a star may be taken as infinitely distant, but it is not always convenient to use such an object. For most purposes it is sufficiently accurate to make use of a terrestrial object, such as a distant church steeple; if the system under test has a long focal length it may be necessary to make a correction for the finite distance of the object.[5]

It is more usual, however, to make use of a collimator in order to obtain parallel light, a suitable object being placed accurately at the focus of the collimator. The image plane of the system is then determined by focussing the observing microscope on the image of this object. This process may be reversed by moving an object, placed in the neighbourhood of the focal plane of the system, until its image coincides with the focal plane of an observing telescope which has been previously focussed for infinity. Another alternative is to make use of an auto-collimating method; a simple way in which this may be applied is illustrated diagrammatically in *Fig.* 1. One half of a scale S (preferably ruled on silvered or platinised glass) is illuminated from the side by a lamp L with the aid of a small right-angle prism P. If the scale is in the focal plane of the objective O, the image of the

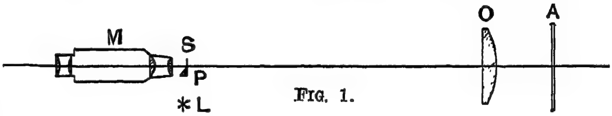


Fig. 1.

[1] The sizes of these portions should be chosen sufficiently large to prevent the definition being spoiled by diffraction effects.

[2] J. Hartmann, *Zeits. Instrumentenk.*, 1900, xx. 51.

[3] L. Foucault, *Ann. l'Observatoire de Paris*, 1859, v. 197; *Recueil des travaux scientif.* Paris, 1878, p. 232.

[4] Cf. the Abbe test plate, for a description of which see *Microscopy*, by E. J. Spitta.

[5] If f is the focal length and v is the distance between the second nodal point and the image plane, the object being at a distance u from the system, then

$$v-f=\frac{f^2}{u}$$

provided that the difference $v-f$ is small.

illuminated portion, formed by light which is reflected back from an optically plane surface A placed normal to the axis, will coincide with the image of the other portion of the scale as seen in the observing microscope M.

These methods only serve to determine the position of the second focal plane and the back focal length (the distance between the last surface and the focal plane) of the system. In order to obtain the focal length one may employ two methods. The first is a direct one; the objective is mounted in such a way that it can be rotated about a fixed axis normal to the optical axis. The whole system is then moved along the optical axis until there is no lateral displacement of the image of a distant object when the objective is rotated about the normal axis. When this is the case, the normal axis passes through a point, called the nul point, which divides the distance between the nodal points into parts whose ratio is equal to the lateral magnification. If the incident light is parallel, the nul point coincides with the second nodal point; the focal length is then given by the distance between the normal axis and the focal plane. The principle of this method is made use of in the testing of camera lenses on the Beck bench.[1]

The second method is an indirect one and depends on the fact that the ratio of the back focal length to the focal length is equal to the ratio of the size of a mark placed on the last surface to the size of its image. This relation follows from equation (2) above, for if f_b is the back focal length, the equation

$$x' = -mf \text{ reduces to } \frac{f_b}{f} = m = \frac{y'}{y},$$

where y' is the size of the mark and y the size of its image. It is sufficient, therefore, to measure the dimensions of the mark as seen direct and through the system. The focal length can then be found if the back focal length is known.

A simple method of measuring the focal length of a photographic lens which can be divided into two parts, each capable of producing a real image of a distant object, is as follows.[2] Let f, f' be the focal lengths of the two components and F that of the complete lens. An image of a distant object is focussed on the ground-glass screen of the camera for four different cases: (1) the complete lens placed normally in the camera, (2) the front component removed, (3) the complete lens reversed in camera, and (4) the component now at the front removed. If d is the distance between the positions of the screen for (1) and (2) and d' the distance between the positions for (3) and (4), then

$$d = \frac{fF}{f'} \text{ and } d' = \frac{f'F}{f}.$$

Hence $\quad F = \sqrt{dd'}$, also $\dfrac{f}{f'} = \sqrt{\dfrac{d}{d'}}$.

These relations follow from the fact that the positions in which the images are formed by the separate components are conjugate foci for the complete lens, namely, that pair of conjugate foci for which the beam of light between the two components is parallel.

(*b*) A method of finding the focal length of a system follows at once from equation (1) above, for according to this equation the focal length is equal to the size of the image (in the second focal plane) of an object at infinity divided by the tangent of the angle subtended by the object at the first focal point. The focal length may thus be obtained by measuring the size of the image and the apparent angular size of the object. The reverse process may also be employed; in this case the apparent angular size of the image of an object of known length, placed in the focal plane, is measured. This method is applicable to telescope, camera, and microscope objectives. In the case of short focus objectives an object of known length y at a finite distance D may be used; the focal length f is then given by

$$f = \frac{y'}{y/D} = \frac{y'}{y}D,$$

where y' is the length of the image. This method can conveniently be carried out on an ordinary microscope supplied with a graduated scale in the focal plane of the eye-piece.

(ii.) *Methods depending on the Properties of Conjugate Points.*—When the positions of the first and second foci of an objective are not known, the focal length may be determined by finding the positions of a number of pairs of axial conjugate points.

(*a*) If the distances of conjugate points P, P′ (*Fig.* 2) from some fixed point Q, such as a mark on the mount of the objective O, are denoted by ξ, ξ', and the distances of Q from the foci are denoted by η, η' respectively, it follows from equation (3) above that

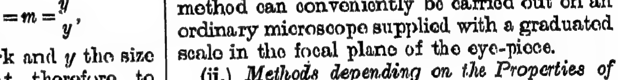


Fig. 2.

$$xx' = (\xi - \eta)(\xi' - \eta') = -f^2.$$

Now this equation contains three unknowns,

[1] See article on "Camera Lenses, The Testing of."
[2] T. Smith, *Phys. Soc. Proc.*, 1915, xxvii. 171.

η, η' and f. Thus the focal length, and the positions of the foci, may be found by measuring the distances ξ_1, ξ_1'; ξ_2, ξ_2'; ξ_3, ξ_3' of three pairs of conjugate points from a fixed point.

(*b*) If N, N′ are the nodal points of the objective,

$$\eta' - \eta = 2f + d,$$

where $d = \mathrm{NN}'$. Thus in cases where d is known or may be neglected it is sufficient to determine the distances ξ_1, ξ_1'; ξ_2, ξ_2' of two pairs of conjugate points. This method is most easily carried out by keeping the positions of the object and image planes fixed and moving O along the axis into the two positions for which the image falls in the latter plane. The distances of object and image from Q are then measured for the two positions, Q being kept fixed relatively to O.

(*c*) The distance D between conjugate points is given by

$$\mathrm{D} = -x + x' + 2f + d.$$

Now it follows from the equation $xx' = -f^2$ that $-x + x'$ is stationary when $-x = x' = \pm f$. The negative sign gives the two nodal points, while the positive sign gives two points on the opposite sides of the foci from the nodal points and at distances from the foci equal to f. In this case

$$\mathrm{D} = 4f + d.$$

Thus if d is known or may be neglected, the focal length may be found by moving the objective until the distance between object and image is a minimum.

It may be noted that in method (ii.) (*a*) the focal length may be found by measuring ξ, ξ' for one pair of conjugate points if the positions of the foci F, F′ are known, that is, if η, η' are known. A knowledge of f then enables one to determine the distance d between the nodal points by means of methods (ii.) (*b*) or (*c*).

The methods in group (ii.) can be carried out most conveniently on an optical bench. They are specially applicable to objectives of medium focal length. The chief objection to employing them in the case of long focus objectives is the great length of space required, the minimum distance between conjugate points being, as we have seen, approximately four times the focal length.

(iii.) *Magnification Methods.*—There are a number of methods of determining focal length which depend on the measurement of the lateral magnification. The most important of these are as follows.

(*a*) The magnification is measured for two positions of the object, the system under test being kept fixed. Then, if x_1, x_2 are the distances of the object from the first principal focal plane and m_1, m_2 are the magnifications for the two positions, it follows from equation (2) above that

$$x_1 = \frac{1}{m_1} f, \quad x_2 = \frac{1}{m_2} f.$$

Thus $$f = \frac{x_1 - x_2}{1/m_1 - 1/m_2}.$$

It is only necessary, therefore, to measure the distance $x_1 - x_2$ through which the object has been moved, in addition to the magnifications. The focal length may also be found by keeping the system fixed, moving the image plane a distance $x_1' - x_2'$, and measuring the corresponding magnifications. In this case

$$x_1' = -m_1 f, \quad x_2' = -m_2 f.$$

Thus $$f = \frac{x_1' - x_2'}{m_2 - m_1}.$$

A simple direct reading method based on this relation can be employed in finding the focal length of a camera lens. The image of a very distant object is focussed on the ground-glass screen of the camera and the screen is then racked back through a distance d until the image of a scale, mounted near the lens in a plane at right angles to the optical axis, is focussed on it. The length of the portion of the scale, whose image lies between two marks separated by a distance d on the screen, is then the focal length of the lens.

(*b*) The distance ξ_1 of the object from a fixed point Q (*Fig.* 2) on the objective is measured and the objective is then reversed and moved along the optical axis until the same magnification is obtained, the object and image planes being thus interchanged. The distance of the original object from Q is now measured. Since the positions of F and F′ are interchanged owing to the reversal of the system, it follows that

$$\xi_1 = -\mathrm{D} + x' + \eta', \quad \xi_2 = x - \eta'.$$

Therefore

$$\xi_1 + \xi_2 = -\mathrm{D} + x + x' = -\mathrm{D} + \frac{1}{m} f - mf$$

or $$f = -\frac{\mathrm{D} + (\xi_1 + \xi_2)}{m - 1/m}.$$

If, in the second position of the system, the distance ξ_2' of the original image plane from Q is measured, then

$$f = -\frac{\xi_1 + \xi_2'}{m - 1/m}.$$

(*c*) The magnifications m_1 and m_2 are measured for two values D_1 and D_2 of the distance between the object and image planes. If, for example, the distance is increased by an amount d, we have

$$\mathrm{D}_2 - \mathrm{D}_1 = d = (-x_2 + x_2') - (-x_1 + x_1'),$$

since $D=-x+x'+FF'$ and FF' remains constant. Thus

$$d=\left(-\frac{1}{m_2}f-m_2f\right)-\left(-\frac{1}{m_1}f-m_1f\right)$$

$$=f\left\{(m_1-m_2)+\left(\frac{1}{m_1}-\frac{1}{m_2}\right)\right\}$$

or
$$f=\frac{dm_1m_2}{(m_1-m_2)(m_1m_2-1)}.$$

(*d*) The system is kept fixed and the positions of the image are determined for which the magnifications are -1, -2, -3, etc. The corresponding distances of the images from the second focal plane are then given by

$$x_1'=f,\quad x_2'=2f,\ x_3'=3f,\ \text{etc}$$

Thus the image has to be moved through a distance equal to the focal length in passing from one magnification to the next. Similarly the distances of the object from the first focal plane, corresponding to the magnifications -1, $-\frac{1}{2}$, $-\frac{1}{3}$, etc., in the image plane, are given by $x_1=-f$, $x_2=-2f$, $x_3=-3f$, etc. In this case, therefore, the object is moved through a distance equal to the focal length in passing from one magnification to the next.

It is most convenient to make use of an optical bench in connection with the methods of group (iii.). The methods, like those of group (ii.), are especially applicable to the case of objectives of medium focal length.

(iv.) *Methods applicable to the Measurement of Short Focal Lengths.*—There are two simple microscope methods which can be used in the case of microscope objectives and eye-pieces; they really come under the previous groups, but may for convenience be dealt with separately.

(*a*) A short focus collimator, having a suitable scale in its focal plane, is fitted under the stage of a microscope, and the objective or eye-piece under test is placed on the stage with its axis collinear with the collimator and microscope axes. The microscope, fitted with a micrometer eye-piece, is then focussed on the image of the collimator scale. If f_1, f_2, f_3 are the focal lengths of the collimator objective, the system under test, and the microscope objective respectively, the magnification observed in the lower focal plane of the micrometer eye-piece is given by

$$\frac{f_2}{f_1}\cdot\frac{\Delta}{f_3},$$

where Δ is the distance between the upper focal plane of the microscope objective and the micrometer scale. The method may be made a direct reading one by choosing the values of f_1, f_3, and Δ, so that $\Delta=f_1f_3$. The magnification observed is then numerically equal to the focal length of the system under test.

(*b*) The objective or eye-piece to be tested is fitted to the end of a microscope tube and a suitable scale is placed on the microscope stage. The magnifications of the scale are then measured for two different tube-lengths by means of a micrometer scale in the eye-piece. The focal length of the system is then given by the difference of the tube-lengths divided by the difference of the magnifications. This follows at once from the relation

$$f=\frac{x_1'-x_2'}{m_2-m_1}.$$

The Abbe focometer[1] is an instrument which is specially designed for carrying out such tests.

§ (3) Measurement of Aberrations.—In designing compound objectives it is not possible to obtain systems which are perfectly corrected; there are always certain amounts of residual aberrations present. In a well-corrected system these are usually very small, so that it is necessary to employ very sensitive methods for measuring them. The chief quantities which require to be determined are spherical and chromatic aberrations, coma, distortion, and curvature of field. In the case of telescope objectives it is mainly the first two which are of primary importance; the measurement of the other quantities is dealt with in the article on "Camera Lenses, The Testing of."

Practically all the methods of measuring the aberrations of an objective depend on the use of an accurately parallel beam of light. If a collimator is employed for this purpose it is necessary that the collimator objective itself should be as free from aberrations as possible, otherwise the results will be vitiated. It is usual, therefore, to employ a well-corrected collimator objective of larger aperture than that of the objective under test. This, however, is not always possible, especially when it is necessary to test large aperture lenses; in such cases one or other of the following methods may be used.

(*a*) When one is dealing with an astronomical telescope objective, it may be fitted on a telescope mount and the measurements made with a star as the object. This gets rid of the necessity of using a collimator, but the method can only be used where special equipment is available.

(*b*) If the objective has not a very long focal length, an illuminated pin-hole mounted at a considerable distance may be used as the object. In this case, however, only approximate values of the aberrations can be obtained owing to the fact that one is not dealing with strictly parallel light, the process of introducing corrections for the finite distance of the object being rather laborious.

[1] S. Czapski, *Zeits. Instrumentenk.*, 1892, xii. 185; see also C. v. Hofe, *Zeits. Techn. Phys.*, 1920, i. 191.

(*c*) Where an objective of similar construction to that of the one under test is available, it may be used as a collimator objective. Then, if the aberrations of each objective are of the same sign and of about the same dimensions, the values of the measured aberrations will be approximately double those of each objective taken separately.

(*d*) If two additional objectives of similar dimensions are available, the three objectives may be tested in pairs, each one in turn serving as the collimator objective. The aberrations of each may then be deduced from the measured aberrations of the combinations, even if the former differ considerably in the three cases.

(*e*) In the case of large objectives, when a suitable collimator objective is not available, a convenient method is one which is now being used at the Reichsanstalt.[1]

The principle of the method is illustrated in *Fig.* 3, which represents a horizontal section of the arrangement. A well-corrected collimator objective C of relatively small aperture, having an illuminated pin-hole S at its focus, is set up with its axis in the plane of, and at right angles to, the axis of the objective O which is to be tested. A good quality pentagonal prism P is mounted on a slide in front of the collimator objective so that it can be moved along the collimator axis across a diameter of the objective O into positions such as P′, P″. Since the direction of the rays which emerge from P is not altered by a slight rotation of the prism about a vertical axis, the rays will be parallel to the objective axis in all positions. Any tilt of the prism about a horizontal axis may be prevented by using a level on the top of the prism and adjusting the prism, if necessary, in its different positions. The method allows one to determine the position of focus for different portions of the aperture. If it is necessary to measure the aberrations along different diameters, the objective may be mounted so that it can be rotated about its axis.

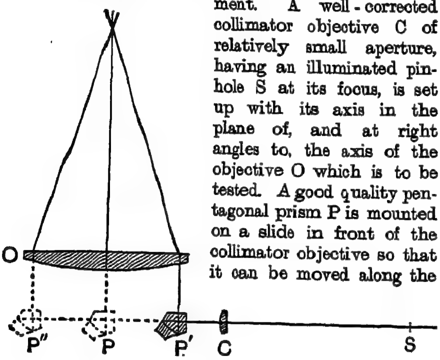


FIG. 3.

(*f*) An auto-collimating method (cf. *Fig.* 1) may be employed, if a sufficiently large optically plane surface is available. This has the advantage of double sensitivity, but for very accurate work the mirror surface used must be plane to a high degree of accuracy.

There are three main methods of measuring the aberrations of an objective. In considering these it will be assumed, as already indicated, that a well-corrected collimator objective of larger aperture than that of the objective under test is employed, though any of the above-mentioned devices may be used instead.

[1] *Zeits. Instrumentenk.*, 1920, xl. 96.

(i.) *Microscope Method.*—The objective, the collimator, and a suitable observing microscope are set up in such a way that their optical axes are collinear, the microscope being placed behind the focal plane of the objective and mounted on a slide so that it can be moved along the axis of the system. A minute pin-hole[2] is fixed at the focus of the collimator objective and is illuminated by a strong source of light. The image of the pin-hole which is formed in the focal plane of the objective consists of a bright circular disc surrounded by a series of concentric diffraction rings. The relative number of rings and the colour visible in planes at small equal distances inside and outside the focus give a qualitative estimate of the spherical and chromatic corrections of the objective.[3] In order, however, to determine these aberrations quantitatively, it is necessary to find the positions of best focus of the central disc when different zones of the objective and different wave-lengths of light are used. For spherical aberration determinations the aperture of the objective should be divided into about 4 or 5 concentric zones of equal area, and a series of opaque stops made of such sizes and shapes that each one only allows light to pass through one of these zones. If the slide, on which the microscope is mounted, is provided with a scale and vernier, or, preferably, an accurate micrometer motion, the axial differences of focus for the different zones can be measured.

In order to measure the chromatic aberration a similar procedure is adopted, colour filters being placed, one at a time, between the pin-hole and the source of light or between the eye-piece of the microscope and the observer's eye. The positions of the microscope, for which the image of the pin-hole is focussed for each of the filters, then enable one to determine the positions of the foci for different colours and hence the chromatic correction of the objective. It is advisable to choose colour filters which are approximately monochromatic and to determine the predominant wave-length transmitted by each. A series of Kodak colour filters, giving about 5 wave-lengths fairly evenly spaced along the spectrum between red and violet, is found very suitable. For very accurate work, however, colour filters may be dispensed with and a monochromatic illuminator placed between the source of light and the collimator pin-hole. In this way monochromatic light of any given wave-length may be employed.

The chromatic differences of spherical

[2] A brightly illuminated small bicycle ball forms a very good substitute for an illuminated pin-hole. In some cases it is preferable to employ fine cross lines or a suitable scale.

[3] See *The Adjustment and Testing of Telescopic Objectives*, by Messrs. T. Cooke & Sons, Ltd., of York.

aberration may be determined by placing concentric stops in front of the objective and measuring the positions of focus for the different zones and the different wave-lengths.

(ii.) *Shadow Method.*—This method, to which reference has already been made in dealing with the methods of finding position of focus, was introduced by Foucault [1] in connection with the figuring of optical surfaces. In order to apply it to the measurement of aberrations a narrow illuminated slit is placed in the focal plane of the collimator objective and a diaphragm with a straight edge is mounted behind the objective under test. The diaphragm is adjusted until its edge is parallel to the slit, the plane of the diaphragm being normal to the axis. If the aperture is covered except for one particular zone, and the observer's eye is placed near the focus of the objective, this zone will be seen illuminated. If, now, the diaphragm is moved across the optical axis in front of the observer's eye, the zone becomes darkened gradually, unless the diaphragm is in the focal plane corresponding to the zone, in which case it becomes darkened suddenly and uniformly. Thus by a method of trial the positions of focus for different zones and different wave-lengths may be found. From these the spherical and chromatic aberrations of the objective may be deduced.

The writer has employed a more symmetrical way of applying the shadow method to the testing of objectives. A small illuminated pin-hole is fixed at the focus of the collimator objective and, instead of the diaphragm with the straight edge, a small opaque circular disc on a glass plate is used. It is convenient to employ a small point image on a photographic plate, its size being approximately that of the diffraction image of the illuminated pin-hole. By means of this arrangement it is possible to find the three co-ordinates of the focus for any zone or to combine the shadow method with the following method (Hartmann's) by measuring the co-ordinates of the points where the pencil of rays from any given region of the objective aperture cuts two planes, one inside and the other outside the focus; in this latter case the plate with the opaque point must be mounted on a mechanism capable of recording measurements in three mutually perpendicular directions.

Methods (i.) and (ii.) are, of course, only applicable to the case of visual measurements. Where it is necessary to determine the aberrations of an objective designed for photographic work, the chromatic corrections being made for wave-lengths in the actinic region of the spectrum, the following method is practically the only one that can be employed; it can be used in the case of any large aperture objective.

[1] L. Foucault, *Ann. de l'Observatoire de Paris*, 1859, v. 197; *Recueil des travaux scientif.* Paris, 1878, p. 232.

(iii.) *Hartmann's Method.*—In this method [2] the positions of the foci for different zones of the objective are determined indirectly. An illuminated pin-hole is mounted axially in the focal plane of the collimator objective, and an opaque diaphragm, containing a number of circular holes arranged in concentric rings, is placed symmetrically in front of the objective. A photographic plate is mounted at right angles to the optical axis in a plane at some distance outside the focus and an exposure made. The plate holder is then moved to a position inside the focus and another plate exposed. Thus records are obtained of the diffusion circles in which the narrow pencils of light, passing through the holes in the diaphragm, meet the two planes in which the plates were placed. The co-ordinates of the centres of these diffusion circles can then be obtained by measuring the plates,[3] and if the distances of the two planes from some fixed arbitrary point have been carefully measured, the positions of the foci, relative to that point, for different zones may be deduced from the relation (4) given above. Chromatic aberrations may be obtained by making a series of measurements with light of different wave-lengths.

If carefully carried out, this method is a very accurate one, but it suffers from the disadvantage that a considerable amount of time is required in order to complete a series of measurements.

In addition to the above methods the aberrations of an objective may be measured by means of interferometry.[4]

§ (4) Determination of General Definition.—The general definition given by an objective depends on the degree to which the system is corrected for the various aberrations. It is difficult, however, to deduce from measurements of the residual aberrations how good the definition will be, for, generally speaking, there are no data available which give the maximum amounts of the different aberrations that may be present without spoiling the definition. A good idea of the definition may, however, be obtained by placing a suitable test object (a fine graduated scale or a microscopic test object may conveniently be employed) at the focus of the collimator objective and examining its image in the focal plane of the objective under test by means of a microscope. As definition is partly a subjective quality, it is always advisable to compare the definition of any

[2] J. Hartmann, *Zeits. Instrumentenk.*, 1900, xx. 51.

[3] Instead of making photographic records the co-ordinates may be determined by means of direct visual observations.

[4] See article on "Interferometers: Technical Applications"; also E. Waetzmann, *Ann. d. Physik.*, 1912, xxxix. 1042; F. Twyman, *Brit. Journ. of Phot.*, 1918, lxv. 556, 567.

given objective with that of a similar objective whose general optical performance is known. The definition of a telescope objective may also be tested by an examination of the extrafocal diffraction images of a distant point source.[1] The resolving power may be determined by observations on double stars.

J. S. A.

OBJECTS, MOUNTING OF, for microscopy with ultra-violet light. See "Microscopy with Ultra-violet Light," § (4).

OBOE: a wind-instrument, played with a small double-cone reed and having a conical tube. See "Sound," § (35).

"ONDE DE CHOC." See "Sound Ranging," § (4).

OPAL GLASSES, MANUFACTURE OF. See "Glass," § (35).

OPHTHALMIC DYNAMOMETER, LANDOLT'S: an ophthalmic instrument for determining the maximum of convergence of a subject's eyes. See "Ophthalmic Optical Apparatus," § (3).

OPHTHALMIC OPTICAL APPARATUS

§ (1).—Until the introduction of cylindrical lenses for spectacle use, ophthalmic optical apparatus was of the crudest, and consisted mainly of sets of pairs of spherical lenses fitted into horn frames known as "triers." With cylindrical lenses and the necessity for rotation of the lenses, came the full trial-case, the ophthalmoscope, ophthalmometer, and retinoscope.

Ophthalmic instruments may be divided roughly into a subjective class, where the results are obtained from the answers of the subject or patient relating his impressions of what he actually sees by means of the instrument, and into an objective class, where results are arrived at by deductions being made from what is seen by the observer in or of the eye of the patient without reference to any questioning of the patient.

§ (2) SUBJECTIVE INSTRUMENTS.—Among subjective instruments we can include:

Phorometers for testing and exercising the strength, deviation, and direction of the external ocular muscles.

Perimeters and scotometers for plotting and measuring the field of vision.

Photometers for light perception.

Chromo-optometers for colour perception.

Ophthalmic lenses.

Test-cases of trial lenses (the Donders method).

Test-types for measurement of visual acuity and form perception.

Diffusion-area instruments relying upon the distortion or aberration produced by ametropia.

Optometers, a term given to instruments generally of telescope design, and dependent upon a personal adjustment of focus by the patient.

§ (3) PHOROMETERS.—Phorometric instruments deal with the measurement of ocular muscular want of balance, and the muscular exercises for its correction. The muscles of a normal pair of eyes are so balanced that the eyes involuntarily converge upon any fixed object, each eye conveying to the brain a similar picture. The exact superposition of these two pictures creates a single picture with stereoscopic relief. Should there be some difference in the size or definition of these two pictures due to difference in the focal strength of the eyes, or should a single picture be obtained with difficulty, the result is either that a strain is thrown upon the extra-ocular muscles in an attempt to secure single stereoscopic vision, or else failure to secure such vision results in a more or less pronounced squint or strabismus, with a temporary suppression of vision in one eye, in order to avoid confusion. It is obvious that, if two separate images are seen by the two eyes, a fusion of the images can be obtained by placing in front of one eye a prism, which will cause one of the images to deviate and superpose itself over the other image, and thus secure stereoscopic vision.

Supposing always that the eye refraction has been corrected, and pictures of equal size and definition secured, it is one of the simplest of optical procedures to find a prism which will give the required deviation. The first thing is to make one of the images seem so altered or distorted that the brain no longer attempts to associate it with the other image, and consequently ceases to make any muscular effort to superpose the two images. This distortion is usually effected by glass rods or cones or similar distorting contrivances.

One of the most effective of these is the Maddox set of rods or grooves (*Fig.* 1), which

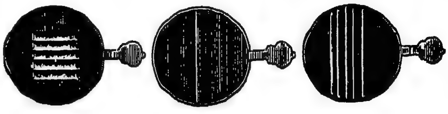


FIG. 1.

consists of a set of short coloured convex cylindrical thin glass rods, or concave cylindrical grooves fixed side by side in a small disc for use in the oculist's trial-frame. A bright small light seen through one of these rods produces a small fine line as an image, the several rods or grooves making in continuation a long decided line of light which can easily be

[1] See *The Adjustment and Testing of Telescopic Objectives*, by Messrs. T. Cooke & Sons, York.

fixed by the patient. This line is so different from the image of the object seen by the other eye that there is no effort on the part of the patient to secure superposition or overcome what is called his "dynamic strabismus." Successive prisms are now placed before one of the eyes until this streak of light appears to coincide in position with the original light.

The Risley or Herschell prism, which consists of two equal-powered prisms superposed and rotating at equal rates in opposite directions, is often used in order to obviate the necessity of trying successive prisms.

Another method of correcting the defect is that of the Stevens phorometer, which consists of two equal prisms (one before each eye) rotating in opposite directions. With either of such instruments, or some similar phorometer, the strength of the external eye muscles may be measured. There are many varieties of this class of instrument, varying chiefly in the number of small trial-case accessories attached to save time in adjustment.

Nearly all instruments for testing the ocular muscles depend upon their ability to create a diplopia which will not be overcome mentally by the patient, and having created it to indicate the prism giving fusion or a single image.

In order to induce diplopia, Thorington uses an ingenious form of truncated cone ground in two 7·00-dioptre prisms, separated by an interval of plane glass 3 mm. wide, which produce the combined effect of a Maddox rod and doubling prism in one piece of glass, with the important addition of a centre light as a starting-point.

An excellent but little-known instrument is the Remy Diploscope, which consists of a wide, short, hollow cylinder blackened on the inside, 28 cm. in length and 9 cm. in diameter, open at the eye-end, and closed at the other end by a disc with four round holes about 2 cm. in diameter, two of which are 4 cm. apart, and the others 1 cm. apart. By revolving a shutter, one pair of holes is kept closed while the other pair is open, and by revolving the disc the pair which is open may be set at an angle which is required for a test. In front of the open end of the cylinder there is a black square frame to block out stray rays of light, and a black bar at the top of the frame can be either lowered vertically across the opening, moved at an angle to one or the other side, or lifted away altogether. This cylinder is mounted on a long rod with its axis parallel to the rod so that the disc containing the holes is midway between the two ends. The length of the rod is 120 cm., the disc is thus 60 cm. from the eye, while the distance between the centres of one pair of holes is approximately equal to the distance between the eyes. At the other end of the rod is a test-card carrying four letters arranged horizontally at distances of about 6 cm. apart. If a patient with binocular vision looks at the holes he will see all four letters. Each eye sees a different letter through each of the two holes. If he suffers from convergent squint only two letters will be visible. By employing the second disc with the holes closer together and a second test-card on which there are two letters one above the other, further tests can be made.

Another popular instrument used in the British army is "Bishop Harman's Diaphragm test." This is the reverse of Javal's well-known bar-reading test. Instead of a bar there is a screen with a single hole in it (*Fig.* 2).

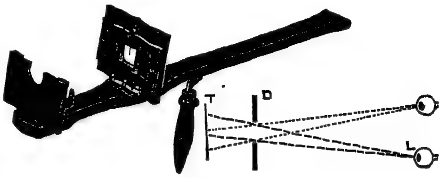


FIG. 2.

Into this hole the patient can look with both eyes without suspecting the test to which his vision is being subjected. The instrument is held with the end of the bar against the face and a screen with a suitable design is mounted in the holder at the other end of the bar. The patient sees part of this screen with the right eye, part with the left, and a small central region with both, as shown in the sketch. From the extent and location of these various fields the nature and amount of the squint, if any, can be deduced.

The difficulty of fitting prisms to cure the defect does not lie in the manipulation of the instruments, but rather in the inability of the patient, as a rule, to wear anything approaching the prism or Maddox correction which gives perfect fusion, possibly due to an inconstant variation in convergence on the part of the patient, and also to the distortion, astigmatism, and chromatic aberration inseparable from prisms of any strength. It is also possible that the exact relation between a prism giving fusion, and the correction of eye-strain and strabismus, has not yet been satisfactorily settled. When this question of fitting prisms and lenses has reached the certainty of determination of that used in the prescription of spherical and cylindrical lenses, a new and universal type of spectacle lens must immediately follow.

In England considerable attention has been given to the question of muscular anomalies. Of late years the tendency has been to recommend muscular exercises for the cure of

strabismus or squint, rather than interference by surgical operation. Nearly all instruments that are used are based, more or less, on the principle of the stereoscope.

For determining the maximum of convergence, an instrument known as Landolt's Ophthalmodynamometer may be used. This consists of a metal cylinder blackened on the outside placed over a small light or candle flame. The cylinder has a vertical slit 0·3 mm. wide covered by ground glass. The luminous vertical line is the object on which attention is to be fixed. Beneath the cylinder is a tape measure graduated in cm. on one side, and on the other in a corresponding number of meter angles. The cylinder is gradually approached towards the patient until double vision of the line of light occurs.

An instrument for testing latent torsion is the Optomymometer of the Geneva Optical Company, which consists of two tubes about 20 in. long, one of which is moved horizontally only, while the other can be elevated or depressed to any angle. At one end of each tube is a rotatable disc in which a slit is cut. On looking down the tubes, one slit is seen with each eye; by adjusting the inclinations of the tubes one slit may be made to appear vertically above the other, and by rotating the discs they may be made to appear parallel. There are many variations of this instrument, one of which is that of the clinoscope of Cooper, and the more recent one of Stevens.

§ (4) Pupilometers.—The size and shape of the pupil and its position in the centre or otherwise in the iris field is a matter of some importance both in discussing pathological questions and also in accounting for the uncertain results sometimes obtained in the estimation of refraction. It has been found that the visual axis rarely passes through the centre of the pupillary opening. Various methods have been devised to measure either the size of the pupil or its position. The simplest, and the one in common use, as in *Fig.* 3, is to have a series of various-sized

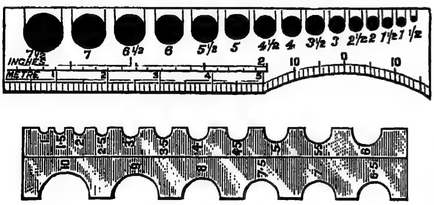


Fig. 3.

black dots on a card, and to compare them with the actual pupil, or to make similar comparisons with an ordinary wedge-shaped gauge. The other methods have been confined to the laboratory, and have consisted chiefly in the use of ophthalmometers with the doubling prisms removed, and by adjusting the size of the images of the mires or bringing them into contact on the cornea until they just cover it, an estimate of the diameter of the pupil has been made. Another ingenious method due to Landolt was to use a Fresnel doubling prism, and so adjust the distance between the two images until they just touch. Photography to scale has also been employed.

§ (5) Perimeters.—Perimeters or scotometers are for the measurement of the field of vision from the macula to the peripheral parts of the retina, and the plotting out of portions of the retina which are totally or partially atrophied. The measurements can be made for the colour as well as the light field.

Roughly the field of vision can be ascertained by making the patient look fixedly at a spot, such as the eye of the surgeon. A test object, such as a small ball, is then placed in front of the eye in various positions until the limits of indirect vision are determined. Usually elaborate instruments are used, allowing the travel of a test object along an arc having the patient's eye as its centre. Such instruments are generally furnished with mechanism for the silent and unobserved movement of the object, and the simultaneous recording of the limits of vision by a puncher, pricker, or pencil on a graphic chart behind the instrument.

Aubert and Forster constructed an instrument in 1847 upon which most of the subsequent perimeters have been based, amongst which are the McHardy, Skeel, Dana, Stevens, Schweigger, Landolt, Bardsley, and Priestley Smith perimeters and scotometers. These vary chiefly in the mechanism for the test-object carrier and the method of making the chart. *Fig.* 4 shows the McHardy instrument.

Tomlinson, in order to avoid the large and cumbersome arc, designed a system of a tilting mirror which gives an imaginary movement to the object.

Many surgeons use the Bjerrum screen, which consists of a large square black screen 2 metres in breadth, at a distance of 2 metres from the patient. This system has the advantage of increasing all the measurements—the blind spot, for instance, measuring 20 cm. on the screen instead of about 2½ cm. with the ordinary perimeter. Beyond 45°, measurements on a flat surface are not of much use. R. H. Elliott has recently made considerable improvements on this method. He uses in front of the Bjerrum background a movable disc in which there are small illuminated objects.

In the Campimeter of De Wecker the patient fixes a small cross in the middle of a black screen on which are marked radial lines;

the test object is moved from the periphery to the centre, the limits of recognition being marked on each line in turn. The points on

FIG. 4.

the various radiating lines are then joined, showing the area of the visual field.

The same principle applies to nearly all the perimeters. In some like the Priestley Smith, the McHardy, the Skeel, etc., the test objects travel on an arc centred round the patient's eye, instead of at a tangent, as in the case of the Wecker and Bjerrum screens. The differences in the many arc instruments are rather in details of workmanship and mechanism than principle.

Hand perimeters are rather for clinical or bedside use.

The uncertainties of perimetry are due chiefly to the tendency of the patient to allow his eye to wander away from the fixation point and also to the various degrees of intelligence on the part of the patient.

§ (6) PHOTOMETERS. — Until the war and recently, comparatively few experiments had been made as to the perception of light in the estimation of refraction and eyesight. The prevalence of night-blindness, the possibilities of malingering, and insurance problems, brought into use some of the many well-known photometers.[1] Perception scales such as those of Parinaud, which consist of ten graded squares of grey colour from white to black or a black background on which various colour shades are outlined in small squares to be counted by the patient, constitute probably the quickest and simplest method.

§ (7) CHROMO-OPTOMETERS. — Since attention has been drawn by the Board of Trade and Admiralty to the necessity for stricter tests, the examination for colour vision has become a much more general practice. Some observers employ the confusion test of a series of coloured wools (Holmgren), while others, such as Butler-Harrison and Edridge-Green, rely upon the use of lanterns showing by rotation and superposition a series of colours which the patient is called upon to classify or compare.

Another useful test is that of the bead-box of Edridge-Green, where a variety of coloured beads is given to the patient to classify.

An ingenious instrument is that of Chibret, in which a plate of quartz ground parallel to its axis is placed between Nicol prisms. The plate is seen coloured, and the colour depends on the thickness of the quartz. If the Nicol analyser is rotated, the colour changes. At 45 degrees the field is white, on rotating the prism to 90 degrees one obtains the complementary colour which increases the intensification of colour. By replacing the analyser with a bi-refringent crystal, in the form of a prism, one of the images has the complementary colour to the other. In the instrument of Chibret, by varying the obliquity of the prism, it is possible to obtain the whole range of colour and intensities. The patient is asked to regulate the instrument so that at any degree of intensity he is sure to secure equal colouring of the two fields. Failure to do this indicates his degree of colour sensitiveness.

§ (8) OPHTHALMIC LENSES. — Ophthalmic spectacle lenses differ from photographic, projection, microscope, or what may be termed image-forming lenses or systems, in that they are usually of approximately one diameter or aperture, one thickness, and except in the cases of special bifocal and cemented lenses, they are thin single lenses of two surfaces only.

Spectacle lenses are made of ordinary white glass similar to that of a good window glass. The English glasses are generally very white but softer than the French or St. Goban glass. Recently Schott of Jena have manufactured a hard and dense glass for better class meniscus lenses.

Until about twenty years ago the more expensive lenses were made of Brazilian pebble or "rock-crystal." It was claimed that the extra density of the crystal allowed a thinner lens, and cooler to the eye; but on the other hand there were many faults, such as that of double refraction, striae, and surface imperfections in working, that made them inferior to ordinary good white glass. The introduction of cylindrical and prism lenses resulted in comparatively few lenses being made of pebble owing to the difficulties and expense of working the surfaces.

[1] See "Photometry and Illumination."

With spectacle lenses a large variety of surfaces and combinations, useful only in ophthalmic lens work, are employed.

(*a*) The plano-spherical, where one side is plane, the other convex or concave spherical.

(*b*) The bi-spherical, where both sides are spherical, either convex-concave or else concave or convex.

(*c*) The plano-cylindrical, where one side is plane and the other is a simple cylinder.

(*d*) The periscopic, a term given in the trade, of which one surface is concave, the other convex, and in which the concave side is about 1·25 to 2 dioptres in strength, the other side having an excess of convexity to secure the necessary focus.

(*e*) Meniscus lenses, practically deep periscopics, where the concave surface ranges from 6·00 to even 12·00 dioptres.

(*f*) Ordinary sphero-cylinders, where one side is convex- or concave-spherical and the other a convex or concave cylinder.

(*g*) Toric lenses, where the one surface is that of a segment of a torus, or anchor ring. A good illustration of a toric surface may be taken from that of a bicycle tyre which has two differing curves, one at right-angles to the other. The advantage of a toric lens is that it provides a sphero-cylindrical lens in the form of a meniscus combination.

Prisms are also used with any of the above lenses or combinations. If the prism required is comparatively weak and the spherical power fairly strong, this may be secured easily by decentration of the lens in the spectacle frame, but if the prism required is fairly strong, without a corresponding increase or ratio of strength in the spherical, then the spherical curves are usually worked upon a prism or wedge-shaped glass.

There are some forms of special spectacle lenses which call for special mention, not only for the special purpose for which they are intended, but on account of their very beautiful workmanship, such as the bifocal, aspherical, and cataract lenses.

(i.) *Bifocal Lenses.*—The bifocal is a spectacle lens having two different powers or foci set in the same eye-wire or spectacle frame; the upper power or focus being used for distance and the lower for reading or close work. A bifocal lens is generally useful in the case of a person who is ametropic and who requires, on account of age, two separate foci, one for distance and one for reading, and wishes them to be in the same spectacle frame in order to avoid having two pairs of spectacles.

There are many forms of bifocals, the oldest being the Franklin bifocal, made of two half lenses of different foci joined together with a straight joining line. The drawback to this pattern is that the half-lenses are liable to jump out of the frame, the junction retains dirt, while the lower half presents inconvenience, and is usually too large for the purpose for which it is intended. This is sometimes avoided by placing the lower half into a curve cut out of the upper piece. Very many suggestions and patents have been taken out for bifocals, but most of them have been discarded on account of the difficulty of securing good centring and also because the dividing line or ridge is generally in the way. The commonest form of bifocal consists of small thin segments of additional powers cemented on to an ordinary lens of the focus required for distance. The segment is usually so thin as to be almost imperceptible, but is liable to become disturbed through heat, strain or concussion. One very modern form is generally known as the fused solid bifocal, which is made by grinding or gouging out a small depression or curve in the lower part of the distance lens, which should be of low refraction. Into this depression or curve or small basin is dropped and fused another small lens of very much denser refraction. The difficulty in manufacture is to secure a perfect join between the two surfaces without introducing air bubbles. After fusing, the whole surface is then ground to a specified curve, in appearance one lens only, although there are two separate foci. The process requires such extreme care in manufacture that the lenses have become almost a proprietary article.

Another equally modern form is that of the solid bifocal, where two separate curves are ground on the same surface. A small but almost imperceptible ridge is formed between the two parts of different foci, which are ground out of the same piece of glass and therefore do not depend on two different refractions. The method of manufacture calls for a kind of optical grinding that has not been used in any other kind of optical work. All opticians are aware that all optical lenses are surfaced by means of optical tools having the same curvature, and by means of emery and rouge are polished afterwards on the same tools with a piece of cloth. This bifocal is ground quite a different way, in that the surfaces of all the lenses, no matter what the foci, are ground by means of a thin ring tube about half the diameter of the part to be ground out from the lens. The end of this tube rotates at a different rate from the lens, which is fixed on a shaft, and the difference in rotation results in varying curves, spherical curves of extra curvature being ground. Theoretically the one tool grinds every curve. The difference in the rate of the abrasion between the outer and inner edges of the cutting tube produces the curve desired.

(ii.) *Aspherical Lenses.*—The optical formulae of ordinary spherical lenses are of the

simplest and do not require discussion. Von Rohr has pointed out that the formulae hold good only in the case when the axis of the spectacle lens is directed towards that particular object-point that is the principal object of contemplation, which would mean that in order to realise such conditions we should have to move our heads continually in looking around: this is avoided in the ordinary way by moving the eye instead of the head and thus getting what is called direct vision.

The result is that when the eye moves and looks towards the edge of the lens, fresh optical conditions are set up which result in aberration. It has not been found possible in actual practice to correct vision by the usual combinations such as are met with in photographic cemented objectives on account of the expense, weight, fragility, and liability of the lenses to become uncemented. Any attempts made to correct this aberration have been in the direction of giving definite form to the surfaces of meniscus lenses, according to the resultant focus required. Wollaston, Ostwalt, Tscherning, Percival, Gullstrand, and Whitwell have at various times drawn up a definite series of tables defining the curves which are likely to give the least amount of aberration with certain foci. The problem has naturally been more difficult in the case of toroidal or cylindrical lenses. Gullstrand has devised an aspherical surface of revolution where the meridian curve is different from the arc of a circle, and includes such surfaces of revolution as elliptic, parabolic, and hyperbolic surfaces. Undoubtedly this is an ideal form of curvature and is limited only by difficulties of manufacture.

(iii.) *Cataract Lenses.*—As lenses for aphakic patients are generally above 13·00 D in strength, the weight of the lens is a great drawback. Attempts have been made to do away with this by cementing on an ordinary plano-lens a small convex lens, which does away with weight and reduces the aperture and consequently the aberration.

§ (9) Test-case of Trial Lenses.—There is probably no piece of ophthalmic apparatus which is more universally used or relied upon by both oculists and opticians than the test-case of trial lenses. No matter what method or instrument may be used, almost invariably, at some stage of the procedure, use is made of the test-case and trial-frame.

The Trial-frame.—Lenses when used for testing a patient's vision are usually mounted in a trial-frame, of which there are numerous patterns with adjustments to suit the patient and to rotate the lenses. *Fig.* 5 illustrates one form.

A test-case consists of a series of the lenses forming the various combinations possible of spectacle lenses. They are nearly always circular, 38 mm. in diameter, unmounted or mounted in metal rings with handles. Some-

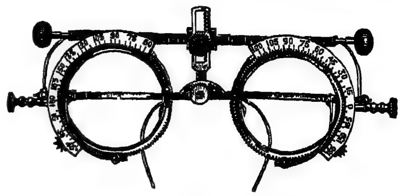


Fig. 5.

times they consist in their simplest form of a small series of about 50 single spherical lenses for use in retinoscopy in the dark room, or they may include a full case of pairs of convex and concave sets of spherical and cylindrical lenses grading by 0·12 intervals up to about 3·00 D, by 0·25 intervals up to about 4·00 D and 0·50 intervals to about 8·00 D. The larger cases generally contain very many accessories, such as a full set of prisms, blanks, pinholes, and stenopaic slit discs, together with a full range of coloured glasses. Some trial-cases also contain extra smaller sets of bifocal and cataract lenses. Almost any combination can be made by combining one, two, or three lenses together, although the distances from the eye and the separation of the lenses must be taken into account when ordering the actual spectacle lenses.

For that reason, the ideal test-case should consist of plano-spherical lenses rather than bi-spherical.

The Optical Society set up some years ago a set of standards for trial-cases, and lenses are tested to these standards at the National Physical Laboratory.

The pinhole discs consist of small various-sized perforated discs which afford a speedy means of determining whether imperfect vision is due to ametropia or to other conditions. In pure ametropia, the pinhole reduces the circles of diffusion and gives good vision with a certain loss of illumination. In other cases it makes the vision worse. The cases in which it is likely to be misleading are of an unequal surfaced or faceted cornea or conical cornea, where better vision is sometimes secured by a pinhole disc and yet no suitable correcting lens can be found.

The stenopaic slit acts very similarly in cases of astigmatism, where the diffusion areas on the retina are greater in one principal meridian than in another. The stenopaic slit limits them at right angles to its length.

The ordinary use of test-lenses is to test each eye separately, giving the strongest convex lens in the case of hypermetropia, and weakest concave one in the case of myopia, consistent in each case with best vision. In similar fashion, but applicable to one meridian

only, use is made of cylindrical lenses for astigmatism. The difficulties likely to be encountered in the subjective use of the trial-case arise chiefly through having to rely upon the answers of the patient and the tendency there is to excite the accommodative apparatus of the eye.

Some operators use, in order to save time, apparatus consisting of a series of test-lenses revolving before the eye of the patient. These are sometimes termed refractometers, or by various proprietary names.

Subjective testing, or the use of trial lenses, cannot be regarded as thoroughly satisfactory until the conditions of the lenses in the trial-frame on the patient's face approach those of the actual spectacles. To this end the lenses should be very thin; if a combination of two is used, then each trial lens should be half the thickness of the spectacle lens. The trial-lens should be plano-spherical or plano-cylindrical. The two plano surfaces of these should face each other; the separation between them should be almost nil; and the distance of the farthest lens from the eye and its tilt should be the same as the actual spectacle lens worn. These conditions of course are difficult to fulfil, although instruments have recently been devised to measure the distance of the spectacle lenses from the eye, and also of the trial lens, and to make comparison.

Many users of test-cases prefer to estimate the refraction by such objective methods as retinoscopy, ophthalmometry, and ophthalmoscopy, putting up an approximate combination into the trial-frame and then by slight adjustment or alteration determining the final correcting glasses. Some, however, prefer to rely entirely on subjective methods and more or less adopt a definite system and rigid routine in its use.

A careful operator who relies principally upon his test-case without preliminary objective findings nearly always uses some kind or other of "fogging" method. This consists in blurring the vision by a positive lens of power in excess of that required for correction and interposing negative lenses of gradually increasing power until distinct vision is obtained. By approaching the desired correction from this side the patient's accommodation is prevented from coming into play and producing an apparent increase in myopia. A fan line diagram or similar test chart should be used, otherwise mixed astigmatism may be overlooked.

One such modern subjective system consists of the use of the usual set of test-letters, a set of fan lines and a block or series of parallel lines capable of being rotated to any particular meridian. The procedure is to use the test-letters only to ascertain the visual acuity before and after testing, and to rely upon fitting separately the two principal meridians without further reference to the test-letters. The patient is asked to state which lines in the fan are the clearest. The set or block of parallel lines is then fixed at right angles to the clearest line and therefore at the worst position at which they can be seen. Convex spherical lenses are tried upon this, and if the vision is improved are increased in strength until the best result is obtained; the fan lines are then again resorted to and the patient is asked to pick out the best set; a suitable concave cylindrical lens is then put in the trial frame and turned until the fan lines appear equal in clearness. Should, however, convex spherical lenses not improve the block lines in the first instance, the procedure is repeated from the start with the block lines at the opposite meridian or axis. If convex sphericals are still useless, concave sphericals are tried and the same procedure followed. If the block is still seen most clearly without the aid of a lens, it is assumed that that particular patient is emmetropic and that a concave cylinder with its axis properly oriented is alone required. The advantage of such a system is that the most difficult complicated cases of astigmatism are gradually eliminated at the outset; there is little tendency to excite the accommodation, and by the use of concave cylindrical lenses only instead of convex the best form of periscopic lenses is secured. Any system which does not allow of an exact placing of the cylinder in the trial frame is haphazard and casual.

§ (10) Diffusion-area Instruments.—These instruments depend upon diffusion areas for the estimation of refraction. The Culbertson prisoptometer in optical construction is rather like an ophthalmometer in that it has a double prism which can be revolved through the various degrees of the scale. The patient looking through the circular opening in the centre of the instrument sees two circles. If, when the instrument is adjusted, the circles are just in contact, the case is one of emmetropia; if they overlap each other it is myopia; and when they separate it is a case of hypermetropia. Theoretically, it is a very ingenious instrument, but as the intelligence of the patient is called into play, exactness and accuracy do not always follow.

Another ingenious diffusion instrument is the ametrometer of Thomson, which consists of two very small flames or sources of light, one stationary and the other movable on a graduated arm, revolving about the first flame as a centre. The method is to move one flame along the arm until the two flames appear to fuse. The approximate strength of the lens required is marked on the scale.

Under the head of different tests may come that of the Cobalt-blue lens. For the purposes

of this test a cobalt-blue glass, which appears dark blue but contains as much red, is fitted in the trial-frame. Cobalt-blue glass has the power to exclude all but blue and red rays. Blue rays are more refrangible, and therefore focus sooner than red. If such a glass is placed before the patient's eye and a small flame used as a test object an emmetropic patient will see a small circle composed of two colours equally mixed—that is to say, purple—but a hypermetropic patient will see a red ring of light with a blue centre, and a myopic patient will see a blue ring with a red centre.

§ (11) Optometers.—Optometry is a term sometimes applied to all ocular methods of estimating the refraction of the eye. We confine the application here to that generally accepted in Europe, *i.e.* to instruments where an adjustment of lenses is made by the patient in order to obtain the clear image of an object and the result recorded on the metric scale. Hence the term. Prior to the introduction of retinoscopy and phorometry, the optometer was almost the only method of estimating the refraction other than the trial case of lenses. There are probably more varieties of optometers than of all other ophthalmic instruments put together, probably due to their construction being so closely allied to that of the ordinary optical bench. In actual practice their use has been almost entirely discontinued, chiefly owing to the uncertain accommodation of the patient which so easily confuses the result. This involuntary act of accommodation is due to the fact that all such optometers are provided with a small eye-piece aperture; and this, combined with the need to rely upon the ability of the patient to decide very accurately as to comparisons of clear focus, calls into play the accommodation and imagination.

The most of the older optometers rely upon the well-known principle of Scheiner's experiment that if a card in which two small holes are pierced at a distance from each other less than the pupillary diameter is held in front of the eye, a luminous point seen from a distance will appear to a normal-sighted person as one light, whereas to one with defective vision, it will appear as two lights.

De la Hire, Porterfield, Thomas Young, Helmholtz, Bull, and many others followed on with modifications of instruments, all based on this principle, but while all of them are optically interesting they are unsatisfactory in actual practice.

Subsequently Coccius, Donders, De Graefe, Perrin, Badal, Hardy, De Zeng, and many others designed optometers to obviate the effects of accommodation. Most of these employ a fixed convex lens serving as an eye-piece and a movable illuminated object. The patient places the object where he sees it most clearly and notes its position relative to the principal focus of the lens. If it be between the lens and the principal focus the rays which enter his eye are divergent, he is short sighted; if it is beyond the principal focus, he is long sighted. The difficulty of obtaining accurate results is that while optically accommodation may be eliminated, yet the mental effect on the patient of seeing something at a distance which really is close to upsets his judgment and he gives inaccurate positions for that of the clearest vision.

Optically the best of these is the instrument of Badal, where a fixed single biconvex lens is placed so that its principal focus coincides either with the nodal point of the patient's eye or with its anterior focus. In either case the size of the retinal image of an object situated at any distance in front of the lens is unaffected by moving the object nearer or further away. This destroys the sense of varying distance and prevents the accommodative efforts which result from this. This optometric system is the best if one can rely upon the accurate judgment of comparison by the patient.

J. H. S.

Optic Axis: a direction in a crystal along which there is no separation of the ordinary and extraordinary rays. Crystals having one such direction are known as uniaxial; those having two such directions as biaxial. See "Polarised Light and its Applications," § (5).

Primary and secondary. See *ibid.* §§ (7) (ii.) and (18) (iii.).

Optical Activity: the power possessed by certain substances of rotating the plane of polarisation of a beam of light passing through them. See "Polarised Light and its Applications," § (20).

Optical Bench, for testing of lenses. See "Camera Lenses, Testing of"; "Lenses, Testing of Simple."

OPTICAL CALCULATIONS

§ (1) Trigonometrical Methods. — The methods employed almost exclusively in the past in the computation of optical systems have consisted in tracing step by step the paths of a few rays selected according to rules which are largely empirical in character. From the point of view of the computer this system has much to recommend it. The amount of calculation involved is limited, and with a certain amount of past experience to indicate in what direction modifications intended for the improvement of a system already approximately determined are most likely to prove satisfactory, the evolution of

systems attaining such a degree of correction as has in the past proved necessary is possible in not too prolonged a time. The methods most extensively employed are trigonometrical, and it will be convenient to record here the systems of equations very commonly used for tracing rays through a series of coaxial spherical refracting surfaces. In the first example the ray lies in a plane passing through the axis of the system, and the problem is simply two-dimensional. This case covers by far the greater part of the calculations normally made. In the second case the ray does not lie in an axial plane, and the portions of the path lying in three successive media will not usually be coplanar.

Suppose that there are n spherical refracting surfaces having their centres of curvature on a straight line coincident with the x axis of co-ordinates. Let the surfaces be distinguished by the n numbers $1, 2, 3, \ldots m, \ldots n$, the order of the numbers being that in which the surfaces are met by a ray of light traversing the system from the object space to the image space. The radius of curvature of a surface is denoted by r, to which is added as a suffix the number of the particular surface. The sign convention adopted, which is almost universally accepted, requires r to be regarded as a positive quantity when the light is incident on the convex side of the surface, negative when incident on the concave side. Thus with a positive radius of curvature, and with the light represented in diagrams passing from the left to the right, the centre of curvature will be situated to the right of the vertex or point in which the refracting segment of the sphere meets the axis. Each surface marks the separation of media of different refractive indices, and it is convenient to use numbers to distinguish these also. The total number of these media is $n+1$, and they will be denoted by the numbers $0, 1, 2, \ldots n$. Thus the number of any medium is the same as that of the surface which bounds it on the left or object side, and is less by unity than the number of the surface forming the boundary on the right or image side. This notation is an easy one to remember, but several others are commonly employed. In some cases even numbers are used for surfaces and odd numbers for media, in others the converse arrangement is adopted, while yet another system involves the introduction of half integers. Such variety in the notation is apt to lead to mistakes, and that here adopted, which is less clumsy than most alternatives, will be generally followed in the articles of this Dictionary.

All the quantities required to specify the position of a ray or the configuration of the refracting system can be divided into two groups, in the first of which the quantity is naturally referable to a particular surface, while in the second it is associated with the medium. In the former group the quantity bears the suffix corresponding to the surface, in the latter case the suffix of the medium is used. Examples of the former are the co-ordinates of the point of refraction at a surface, or the angles of incidence and refraction; and of the latter the refractive index, the axial separation of the two bounding surfaces, and co-ordinates giving the direction of a ray in the medium.

The refractive index throughout the present article will be denoted by μ. The axial distance between the vertices of two surfaces is denoted by t, and will for convenience be referred to as the thickness whether this distance is the actual thickness of a lens or the axial separation of neighbouring surfaces of two lenses. The distance between the centres of curvature of two surfaces may be denoted by a, so that if t is regarded as essentially positive, and a is positive if the centre of curvature of the surface of lower suffix is to the left of that which follows, the relation

$$r_m + a_m = t_m + r_{m+1} \quad . \quad . \quad . \quad (1)$$

will hold in all cases. In trigonometrical calculations the formulae present themselves most readily in a form suited for logarithmic computation when the centre of curvature is taken as a reference point, and the separation of successive centres of curvature must be determined from this equation.

Symbols are required to denote the quantities by which the position of the ray is determined in relation to the refracting surfaces. The angle made by the ray with the axis of symmetry of the system will be denoted by ψ with the addition of the suffix of the medium. The angles of incidence and refraction at any surface will be represented by ϕ and ϕ' respectively, with the surface suffix. When the ray lies in an axial plane the point in which it meets the axis, together with ψ, will fix its position. This point is referred to the centre of curvature of the surface, and consequently the surface suffix is used. The

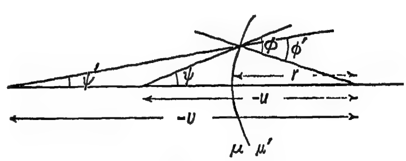


Fig. 1.

distances from the centre of curvature, measured in the light direction, to the crossing points of an incident and the corresponding refracted ray will be denoted by u and v respectively (*Fig.* 1).

For a single surface it is evident that

$$\sin\phi = -\frac{u}{r}\sin\psi \quad . \quad . \quad . \quad (2)$$

is the equation from which ϕ will be found when the position of the incident ray is given. ϕ' is, of course, determined from the refraction relation

$$\mu'\sin\phi' = \mu\sin\phi, \quad . \quad . \quad . \quad (3)$$

and since the distance on the surface of the point where refraction takes place from the axis is given by either $r(\phi-\psi)$ or $r(\phi'-\psi')$, the inclination to the axis of the refracted ray is obtained from the angular relation

$$\psi' = \psi - \phi + \phi'. \quad . \quad . \quad . \quad (4)$$

The remaining quantity v required to specify the position of the refracted ray is derived from

$$-v = \frac{r\sin\phi'}{\sin\psi'}. \quad . \quad . \quad . \quad (5)$$

For a series of surfaces these formulae, together with one additional relation transferring the reference point for the refracted ray to the centre of curvature of the succeeding surface, take the form

$$\sin\phi_m = -\frac{u_m}{r_m}\sin\psi_{m-1},$$

$$\mu_m\sin\phi'_m = \mu_{m-1}\sin\phi_m,$$

$$\psi_m = \psi_{m-1} - \phi_m + \phi'_m,$$

$$-v_m = \frac{r_m\sin\phi'_m}{\sin\psi_m},$$

$$u_{m+1} = r_m - a_m.$$

These equations are obviously well suited for logarithmic work, and the only details requiring special attention are the signs. For paraxial rays[1] the same formulae are generally used by trigonometrical workers with the sines of the angles replaced by any convenient multiple of the angles themselves. Where this system is used tables of logarithmic sines based on the decimal division of the radian are much to be preferred to those in ordinary use.

The chief weakness of this system is that when a radius becomes infinitely great an entirely distinct set of formulae must be used, and that when any radius is relatively large an increase in the number of significant figures is necessary to avoid loss of accuracy. To obviate this latter difficulty special formulae for long radii are frequently introduced. For an account of these reference may be made to *Applied Optics*, Steinheil and Voit, translated by French.

[1] *I.e.* rays always close to the axis at the time of refraction and only slightly inclined thereto.

The process of tracing rays trigonometrically when they do not lie in an axial plane is much more troublesome. The system most generally used is due to von Seidel, and is discussed in the work just mentioned. The formulae may be readily verified by reference to the accompanying figure.

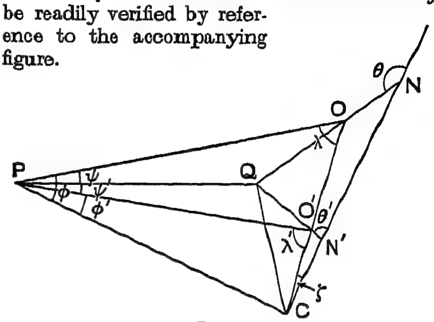


FIG. 2.

The incident ray PO encounters at P a refracting sphere of radius r and centre C. PO′ is the refracted ray. The plane through C normal to the axis cuts these rays in O and O′, and the normal to it through P in Q. An arbitrary reference plane through the axis cuts QO and QO′ in N and N′. C, O, O′ are collinear, and the line determined by them makes angles λ and λ' with the incident and refracted rays. Denote the lengths CO and CO′ by l and l' respectively. Let ζ be the angle made by the lines OO′ and NN′ which meet at C, and let QON, QO′N′ make angles θ and θ' with CNN′. Other letters retain the meanings already applied to them in considering a ray in an axial plane. The formulae to be used suppose that l, ζ, θ, ψ are given and that the corresponding quantities for the next refraction are to be found. They are

$$\cos\lambda = \sin\psi\cos(\theta-\zeta),$$

$$r\sin\phi = l\sin\lambda,$$

$$\mu'\sin\phi' = \mu\sin\phi,$$

$$\lambda' = \lambda + \phi - \phi',$$

$$l' = \frac{r\sin\phi'}{\sin\lambda'} = \frac{l\mu\sin\lambda}{\mu'\sin\lambda'},$$

$$\frac{\sin\psi'\sin(\theta'-\zeta)}{\sin\lambda'} = \frac{\sin\psi\sin(\theta-\zeta)}{\sin\lambda},$$

$$\sin\psi'\cos(\theta'-\zeta) = \cos\lambda'.$$

These give l', θ', ψ'; it is evident that ζ is unchanged. For the next incidence introduce the suffix 1. Then simple projection gives

$$l_1\sin(\theta_1-\zeta_1) = l'\sin(\theta'-\zeta),$$

$$l_1\cos(\theta_1-\zeta_1) = l'\cos(\theta'-\zeta) - a\tan\psi',$$

which completes the solution on noting that

$$\theta_1 = \theta', \quad \psi_1 = \psi'.$$

A number of other distinct solutions have been evolved by various workers, but into these it is unnecessary to enter. This example illustrates fairly the complexity of the best trigonometrical methods of tracing skew rays.

§ (2) Algebraic Methods.—Such methods of calculation as have just been described, essentially trigonometrical in character, were for a period the only means of obtaining results of high accuracy for general optical systems. The recent development of machines for the mechanical performance of arithmetical operations has largely altered the prospects, and at the present time calculations of this kind can be carried out much more expeditiously without employing logarithms at all. The old formulae for the most part seem to be still used with the new tools, but this course is not to be recommended, for greater accuracy and more valuable information become available by an entire change in the basis of the calculations. There is no need with the new formulae to employ a trigonometrical notation, for tables need not be used, though in certain cases they may be introduced as an independent check on a series of operations. It is convenient to have a separate name for these newer methods, and the distinction just noted provides the convenient distinctive name "algebraic" as opposed to "trigonometric" which describes the older methods.

In developing these algebraic methods it is desirable to conform as far as possible with the conventions generally adopted in analytical geometry of three dimensions, and in some branches of mathematical physics. All points and lines will accordingly be referred to a right-handed system of orthogonal axes, of which the axis of x is assumed to coincide with the axis of symmetry of the optical train. The light in general travels in the direction of x increasing. The co-ordinates of a typical point on an incident ray may be denoted by (x, y, z), and those of a point on a refracted ray by (x', y', z'). The direction cosines of a ray may be represented by L, M, N. The point on a surface at which refraction takes place may be represented by (ξ, η, ζ). It is convenient to introduce R, equal to $1/r$, to represent the curvature of the refracting surface. Consider for the moment a system consisting of a single surface. The general problem awaiting solution is the determination of a set of six quantities $(x', y', z', \mathrm{L}', \mathrm{M}', \mathrm{N}')$ for the refracted ray when any set $(x, y, z, \mathrm{L}, \mathrm{M}, \mathrm{N})$ has been given for the incident ray. The origin may conveniently be chosen as the intersection of the axis with the refracting surface, so that the equation to the surface is

$$2\xi=(\xi^2+\eta^2+\zeta^2)\mathrm{R}. \qquad . \quad . \quad (6)$$

Now since (ξ, η, ζ) is on the given ray, the co-ordinates are of the form

$$\left.\begin{aligned}\xi&=x-\mathrm{L}\rho\\ \eta&=y-\mathrm{M}\rho\\ \zeta&=z-\mathrm{N}\rho\end{aligned}\right\}, \qquad . \quad . \quad . \quad . \quad (7)$$

where ρ is the distance along the ray from the surface to the known point in the positive direction of travel. The substitution of these values in the surface equation gives

$$2(x-\mathrm{L}\rho)=\{x^2+y^2+z^2-2\rho(\mathrm{L}x+\mathrm{M}y+\mathrm{N}z)+\rho^2\}\mathrm{R}$$

as the equation from which ρ is to be found. The solution to be chosen is that in which ρ tends to zero with x, y, z: thus

$$\left.\begin{aligned}\rho&=\mathrm{L}(x-r)+\mathrm{M}y+\mathrm{N}z-[\{\mathrm{L}(x-r)+\mathrm{M}y+\mathrm{N}z\}^2-x^2-y^2-z^2+2xr]^{\frac{1}{2}}\\ &=\frac{x^2+y^2+z^2-2xr}{\mathrm{L}(x-r)+\mathrm{M}y+\mathrm{N}z+[\{\mathrm{L}(x-r)+\mathrm{M}y+\mathrm{N}z\}^2-x^2-y^2-z^2+2xr]^{\frac{1}{2}}}\\ &=\frac{2x-(x^2+y^2+z^2)\mathrm{R}}{\mathrm{L}-(\mathrm{L}x+\mathrm{M}y+\mathrm{N}z)\mathrm{R}+[\{\mathrm{L}-(\mathrm{L}x+\mathrm{M}y+\mathrm{N}z)\mathrm{R}\}^2+2x\mathrm{R}-(x^2+y^2+z^2)\mathrm{R}^2]^{\frac{1}{2}}}\end{aligned}\right\}, \quad (8)$$

the last form being always determinate. The point where refraction takes place is thus known. It is next necessary to find the direction of the refracted ray. To obtain this in a suitable form for algebraic calculation consider the triangle PUV (*Fig.* 3), where P is the point of refraction and PU and PV lie

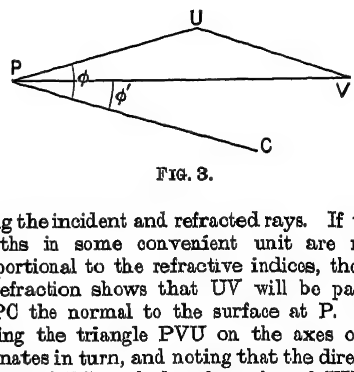


Fig. 3.

along the incident and refracted rays. If these lengths in some convenient unit are made proportional to the refractive indices, the law of refraction shows that UV will be parallel to PC the normal to the surface at P. Projecting the triangle PVU on the axes of co-ordinates in turn, and noting that the direction cosines of PC, and therefore also of UV, are $1-\xi\mathrm{R}$, $-\eta\mathrm{R}$, $-\zeta\mathrm{R}$,

$$\mu'\mathrm{L}'-\mathrm{UV}(1-\xi\mathrm{R})-\mu\mathrm{L}=0,$$

$$\mu'\mathrm{M}'-\mathrm{UV}\ (-\eta\mathrm{R})-\mu\mathrm{M}=0,$$

$$\mu'\mathrm{N}'-\mathrm{UV}\ (-\zeta\mathrm{R})-\mu\mathrm{N}=0,$$

or since $\quad \mathrm{UV}=\mu'\cos\phi'-\mu\cos\phi,$

the refraction equations take the form

$$\frac{\mu'\mathrm{L}'-\mu\mathrm{L}}{1-\xi\mathrm{R}}=\frac{\mu'\mathrm{M}'-\mu\mathrm{M}}{-\eta\mathrm{R}}=\frac{\mu'\mathrm{N}'-\mu\mathrm{N}}{-\zeta\mathrm{R}}$$

$$=\mu'\cos\phi'-\mu\cos\phi. \quad (9)$$

This equation requires ϕ and ϕ' to have been previously determined. The equations for these are

$$\cos\phi = \mathrm{L}-(\mathrm{L}\xi+\mathrm{M}\eta+\mathrm{N}\zeta)\mathrm{R} \quad . \quad (10)$$

and

$$\mu'\sin\phi' = \mu\sin\phi. \quad . \quad . \quad (3)$$

Finally the co-ordinates of any point on the refracted ray are of the form

$$\frac{x'-\xi}{\mathrm{L}'}=\frac{y'-\eta}{\mathrm{M}'}=\frac{z'-\zeta}{\mathrm{N}'}=\rho', \quad . \quad (11)$$

where ρ' is the distance of the point selected from the place of refraction.

Although the equations as they stand provide a complete solution of the problem, they are not in the best form for numerical calculation since an unnecessary amount of work is involved. From (10) and (7)

$$\begin{aligned}\cos\phi &= \mathrm{L}-(\mathrm{L}x+\mathrm{M}y+\mathrm{N}z-\rho)\mathrm{R}\\ &= [\{\mathrm{L}-(\mathrm{L}x+\mathrm{M}y+\mathrm{N}z)\mathrm{R}\}^2+2x\mathrm{R}\\ &\qquad -(x^2+y^2+z^2)\mathrm{R}^2]^{\frac{1}{2}} \quad (12)\end{aligned}$$

by (8), a result more directly obtainable from a figure in the form

$$\begin{aligned}\sin^2\phi &= (1-x\mathrm{R})^2+(y^2+z^2)\mathrm{R}^2\\ &\qquad -\{\mathrm{L}-(\mathrm{L}x+\mathrm{M}y+\mathrm{N}z)\mathrm{R}\}^2. \quad (13)\end{aligned}$$

The use of this equation and the square of (3) enables the last of the equal quantities of (9) to be calculated on the machine by extracting two square roots. Of these, that representing $\cos\phi$ is the square root which appears in the final form in equations (8). The complete solution thus depends upon these two square roots and direct multiplication.

When a series of surfaces with their centres of curvature on the axis of x are substituted for the single surface, the values of ξ, η, ζ for one surface may be taken as the x, y, z, with a correction t for the x co-ordinate, for the next surface. It is thus only necessary to repeat the process described for each surface in turn. The problem may, however, be much simplified by noting that no attention need be paid to the y and z co-ordinates. Suppose that m denotes a typical member of a train of refracting surfaces numbered from 1 to n. Denote by K_m the quantity

$$(\mu_m\cos\phi'_m-\mu_{m-1}\cos\phi_m)\mathrm{R}_m, \quad . \quad (14)$$

and let τ_m be the length of the ray intercepted between surfaces m and $m+1$. Assume for the moment that all the τ's and K's are known. The y and z equations in (9) and (11) take the form

$$\left.\begin{aligned}\mu_m\mathrm{M}_m &= \mu_{m-1}\mathrm{M}_{m-1}-\eta_m\mathrm{K}_m\\ \eta_{m+1} &= \eta_m+\mu_m\mathrm{M}_m\frac{\tau_m}{\mu_m}\\ \text{and}\quad \mu_m\mathrm{N}_m &= \mu_{m-1}\mathrm{N}_{m-1}-\zeta_m\mathrm{K}_m\\ \zeta_{m+1} &= \zeta_m+\mu_m\mathrm{N}_m\frac{\tau_m}{\mu_m}\end{aligned}\right\}, \quad . \quad (15)$$

from which it is evident that linear relations connect any three of the four members of the groups

$$\mu_0\mathrm{M}_0,\ \eta_1,\ \mu_n\mathrm{M}_n,\ \eta_n,$$
$$\mu_0\mathrm{N}_0,\ \zeta_1,\ \mu_n\mathrm{N}_n,\ \zeta_n.$$

the coefficients being linear functions of the various K's and τ's. The forms of these coefficients may easily be found by induction. Construct the quantities

$$\mathrm{K}_{1,n},\ \frac{\partial\mathrm{K}_{1,n}}{\partial\mathrm{K}_1},\ \frac{\partial\mathrm{K}_{1,n}}{\partial\mathrm{K}_n},\ \frac{\partial^2\mathrm{K}_{1,n}}{\partial\mathrm{K}_1\partial\mathrm{K}_n}$$

from the equations

$$\left.\begin{aligned}\mathrm{K}_{1,m} &= \mathrm{K}_{1,m-1}+\mathrm{K}_m\frac{\partial\mathrm{K}_{1,m}}{\partial\mathrm{K}_m}\\ \frac{\partial\mathrm{K}_{1,m}}{\partial\mathrm{K}_1} &= \frac{\partial\mathrm{K}_{1,m-1}}{\partial\mathrm{K}_1}+\mathrm{K}_m\frac{\partial^2\mathrm{K}_{1,m}}{\partial\mathrm{K}_1\partial\mathrm{K}_m}\\ \frac{\partial\mathrm{K}_{1,m}}{\partial\mathrm{K}_m} &= \frac{\partial\mathrm{K}_{1,m-1}}{\partial\mathrm{K}_{m-1}}-\frac{\tau_{m-1}}{\mu_{m-1}}\mathrm{K}_{1,m-1}\\ \frac{\partial^2\mathrm{K}_{1,m}}{\partial\mathrm{K}_1\partial\mathrm{K}_m} &= \frac{\partial^2\mathrm{K}_{1,m-1}}{\partial\mathrm{K}_1\partial\mathrm{K}_{m-1}}-\frac{\tau_{m-1}}{\mu_{m-1}}\frac{\partial\mathrm{K}_{1,m-1}}{\partial\mathrm{K}_1}\end{aligned}\right\}, \quad (16)$$

by putting m in turn equal to 1, 2 . . . n. The initial values are

$$\mathrm{K}_{1,1}=\mathrm{K}_1;\quad \mathrm{K}_{1,0}=0;\quad \frac{\partial\mathrm{K}_{1,1}}{\partial\mathrm{K}_1}=1;\quad \frac{\partial^2\mathrm{K}_{1,1}}{\partial\mathrm{K}_1^2}=0. \quad (17)$$

Since the equations are linear the quantities thus defined by separate symbols justify the differential form selected, which indicates how the four quantities to which the equations lead are related to one another. It is important to note that the four are not independent in value. For, on combining the first and third of the equations,

$$\mathrm{K}_{1,m}=\left(1-\frac{\tau_{m-1}}{\mu_{m-1}}\mathrm{K}_m\right)\mathrm{K}_{1,m-1}+\mathrm{K}_m\frac{\partial\mathrm{K}_{1,m-1}}{\partial\mathrm{K}_{m-1}},$$

and similarly from the second and fourth

$$\frac{\partial\mathrm{K}_{1,m}}{\partial\mathrm{K}_1}=\left(1-\frac{\tau_{m-1}}{\mu_{m-1}}\mathrm{K}_m\right)\frac{\partial\mathrm{K}_{1,m-1}}{\partial\mathrm{K}_1}+\mathrm{K}_m\frac{\partial^2\mathrm{K}_{1,m-1}}{\partial\mathrm{K}_1\partial\mathrm{K}_{m-1}},$$

and therefore on simplification

$$\begin{aligned}&\frac{\partial\mathrm{K}_{1,m}}{\partial\mathrm{K}_1}\cdot\frac{\partial\mathrm{K}_{1,m}}{\partial\mathrm{K}_m}-\mathrm{K}_{1,m}\frac{\partial^2\mathrm{K}_{1,m}}{\partial\mathrm{K}_1\partial\mathrm{K}_m}\\ &\qquad =\frac{\partial\mathrm{K}_{1,m-1}}{\partial\mathrm{K}_1}\cdot\frac{\partial\mathrm{K}_{1,m-1}}{\partial\mathrm{K}_{m-1}}-\mathrm{K}_{1,m-1}\frac{\partial^2\mathrm{K}_{1,m-1}}{\partial\mathrm{K}_1\partial\mathrm{K}_{m-1}},\end{aligned}$$

and is therefore independent of the value of m. On putting $m=1$ the left side reduces at once to unity and thus the relation

$$\frac{\partial\mathrm{K}_{1,n}}{\partial\mathrm{K}_1}\cdot\frac{\partial\mathrm{K}_{1,n}}{\partial\mathrm{K}_n}-\mathrm{K}_{1,n}\frac{\partial^2\mathrm{K}_{1,n}}{\partial\mathrm{K}_1\partial\mathrm{K}_n}=1 \quad . \quad (18)$$

connects the four quantities.

Returning now to equations (15), say the first pair, assume that for some value of m it has been shown that

$$\left.\begin{aligned}\mu_m\mathrm{M}_m &= \mu_0\mathrm{M}_0\frac{\partial\mathrm{K}_{1,m}}{\partial\mathrm{K}_1}-\eta_1\mathrm{K}_{1,m}\\ \text{and}\quad \eta_m &= -\mu_0\mathrm{M}_0\frac{\partial^2\mathrm{K}_{1,m}}{\partial\mathrm{K}_1\partial\mathrm{K}_m}+\eta_1\frac{\partial\mathrm{K}_{1,m}}{\partial\mathrm{K}_m}\end{aligned}\right\}, \quad . \quad (19)$$

both of which are evidently true for $m=1$. From the second of equations (15)

$$\tau_{m+1} = -\mu_0 M_0 \left\{ \frac{\partial^2 K_{1,m}}{\partial K_1 \partial K_m} - \frac{\tau_m}{\mu_m}\frac{\partial K_{1,m}}{\partial K_1} \right\} + \eta_1 \left\{ \frac{\partial K_{1,m}}{\partial K_m} - \frac{\tau_m}{\mu_m} K_{1,m} \right\}$$

$$= -\mu_0 M_0 \frac{\partial^2 K_{1,m+1}}{\partial K_1 \partial K_{m+1}} + \eta_1 \frac{\partial K_{1,m+1}}{\partial K_{m+1}},$$

by (16). Hence also

$$\mu_{m+1} M_{m+1} = \mu_0 M_0 \left\{ \frac{\partial K_{1,m}}{\partial K_1} + K_{m+1} \frac{\partial^2 K_{1,m+1}}{\partial K_1 \partial K_{m+1}} \right\} - \eta_1 \left\{ K_{1,m} + K_{m+1} \frac{\partial K_{1,m+1}}{\partial K_{m+1}} \right\}$$

$$= \mu_0 M_0 \frac{\partial K_{1,m+1}}{\partial K_1} - \eta_1 K_{1,m+1}$$

by the same equations, thus verifying the accuracy of the forms assumed in (19). Evidently another relation of the same form holds if N and ζ are substituted for M and η respectively. The calculation of the four quantities thus defined therefore serves to determine the direction of the emergent ray, since two direction cosines are found, and the position of the point in which the ray meets the last surface, since this is known from two of the co-ordinates. The ray is thus completely determined.

Before considering how the K's and τ's are to be found, it is desirable to note the symmetrical character of the K's found from formula (16). If the ray already considered were retraced through the system in the reverse direction the individual K's and τ's would be unaltered. This is evident as regards the latter, since they represent lengths which are essentially positive. As regards the former the first factor changes sign since the various refracting media are encountered in the reverse order; this change of sign is compensated by a change in the sign of the curvature, for the surface which was before convex to the incident light is now concave, and *vice versa*. The individual powers, like the individual τ's, thus remain unchanged. It follows that the various quantities to be derived on retracing the path of the ray will be identical with those already obtained if $K_{1,n}$ is symmetrically composed of the component K's and τ's with respect to the two ends. When n is small it is easy to see that this is the case. For instance, if $n=2$,

$$K_{1,2} = K_1 + K_2 - \frac{\tau_1}{\mu_1} K_1 K_2,$$

and if $n=3$,

$$K_{1,3} = K_1 + K_2 + K_3 - \frac{\tau_1}{u_1} K_1 K_2 - \left(\frac{\tau_1}{\mu_1} + \frac{\tau_2}{\mu_2} \right) K_1 K_3 - \frac{\tau_2}{\mu_2} K_2 K_3 + \frac{\tau_1}{\mu_1} \frac{\tau_2}{\mu_2} K_1 K_2 K_3$$

when written out in full. To show it in general construct the quantities

$$K'_{1,n}, \quad \frac{\partial K'_{1,n}}{\partial K_1}, \quad \frac{\partial K'_{1,n}}{\partial K_n}, \quad \frac{\partial^2 K'_{1,n}}{\partial K_1 \partial K_n}$$

by repeated application of

$$\left. \begin{aligned} K'_{m,n} &= K'_{m+1,n} + K_m \frac{\partial K'_{m,n}}{\partial K_m} \\ \frac{\partial K'_{m,n}}{\partial K_m} &= \frac{\partial K'_{m+1,n}}{\partial K_{m+1}} - \frac{\tau_m}{\mu_m} K'_{m+1,n} \\ \frac{\partial K'_{m,n}}{\partial K_n} &= \frac{\partial K'_{m+1,n}}{\partial K_n} + K_m \frac{\partial^2 K'_{m,n}}{\partial K_m \partial K_n} \\ \frac{\partial^2 K'_{m,n}}{\partial K_m \partial K_n} &= \frac{\partial^2 K'_{m+1,n}}{\partial K_{m+1} \partial K_n} - \frac{\tau_m}{\mu_m} \frac{\partial K'_{m+1,n}}{\partial K_n} \end{aligned} \right\} . \quad (20)$$

It follows exactly as before, that

$$\frac{\partial K'_{1,n}}{\partial K_1} \cdot \frac{\partial K'_{1,n}}{\partial K_n} - K'_{1,n} \frac{\partial^2 K'_{1,n}}{\partial K_1 \partial K_n} = 1, \quad . \quad (21)$$

and that

$$\left. \begin{aligned} \mu_0 M_0 &= \mu_n M_n \frac{\partial K'_{1,n}}{\partial K_n} + \eta_n K'_{1,n}, \\ \eta_1 &= \mu_n M_n \frac{\partial^2 K'_{1,n}}{\partial K_1 \partial K_n} + \eta_n \frac{\partial K'_{1,n}}{\partial K_1} \end{aligned} \right\}, \quad . \quad (22)$$

as may indeed be inferred by analogy on noting that the M's and N's are changed in sign with the reversal of the ray, while the η's and ζ's are unaltered. Now eliminate η_n from (22); thus

$$\mu_0 M_0 \frac{\partial K'_{1,n}}{\partial K_1} - \eta_1 K'_{1,n} = \mu_n M_n$$

by (21); similarly on eliminating M_n

$$-\mu_0 M_0 \frac{\partial^2 K'_{1,n}}{\partial K_1 \partial K_n} + \eta_1 \frac{\partial K'_{1,n}}{\partial K_n} = \eta_n,$$

with similar relations between the N's and ζ's. Comparison of these relations with (19) shows that $K_{1,n}$ and $K'_{1,n}$ are identical.

It still remains to find the individual K's and τ's so that four co-ordinates sufficient to determine the final ray can be found from (19) and the corresponding equations with N and ζ substituted for M and η. Formulae for this purpose may be arranged in many different forms, each of which has its own special advantage. In general the simplest forms for the purposes of numerical work are open to the objection that they assume indeterminate forms for flat surfaces and cause loss of accuracy for surfaces of slight curvature. On the other hand, forms may readily be found of universal applicability which involve no loss of accuracy, but they require a slight extension in the number of operations at each surface.

It is not possible in the space available for this article to discuss thoroughly any of these forms, but since accuracy is usually more important than brevity one of the many possible arrangements of the second class may be noted. Let the square of the distance of the point of refraction from the vertex of the surface be denoted by 2χ, and the lengths intercepted on the incident and refracted rays between the feet of perpendiculars to them from the vertex and the point of refraction be respectively p and p'. χ is invariant on refraction, and by (6)

$$\xi = \chi R,$$

and therefore from (9) p and p' are connected by the relation

$$\mu'p'=\mu p-\chi K, \quad . \quad . \quad . \quad (23)$$

while the inclination of the ray to the axis after refraction takes the form

$$L'=\cos\phi'+p'R. \quad . \quad . \quad . \quad (24)$$

It is at once evident that to complete the system for tracing rays it is only necessary to derive equations giving the values of χ, p, and $\cos\phi$ for the next surface. Equation (12) gives

$$\sin^2\phi_{m+1}=(1+t_mR_{m+1})^2-\{L_m(1+t_mR_{m+1})-p'_mR_{m+1}\}^2-2\chi_m a_m R_m R^2_{m+1}, \quad (25)$$

a result easily derived directly from a figure. Since

$$a_mR_mR_{m+1}=R_m-R_{m+1}+t_mR_mR_{m+1},$$

every term in the equation is always finite, and $\cos\phi_{m+1}$, $\cos\phi'_{m+1}$, and K_{m+1} are derived in the usual way after the extraction of two square roots. From equation (8) the value of τ_m is found—

$$\tau_m=\frac{t_m(2+t_mR_{m+1})-2\chi_m a_mR_mR_{m+1}}{L_m(1+t_mR_{m+1})-p'_mR_{m+1}+\cos\phi_{m+1}}, \quad (26)$$

and other obvious relations are

$$p_{m+1}=p'_m+\tau_m-t_mL_m, \quad . \quad . \quad (27)$$

$$\chi_{m+1}(2+t_mR_{m+1})=\chi_m(2-t_mR_m)+\tau_m(p_{m+1}+p'_m). \quad (28)$$

§ (3) Advantages of the Method.—The whole process, though quite straightforward, is somewhat tedious, but experience shows that in this respect it is in no way inferior to the logarithmic method generally used in the past, for the references to tables which that system involves, and which take up a large part of the total time, are completely avoided. The advantages of the algebraic method do not, however, end here, for the results not only give the point of intersection of the emergent ray with the assumed image surface, but it is seen at once whether the whole of the image in the neighbourhood of the point to which the rays relate is free from defects. In the most usual case both object and image surfaces are required to be planes normal to the axis. If aberrations are absent in the image of the point from which the rays considered have been traced, the condition for freedom in the neighbouring portion of the image plane is that the values of $K_{1,n}$ for all these rays shall be identical. Further important information is derivable from the way in which the value of $K_{1,n}$ depends upon the position of the object point from which the ray has been traced.

The significance of $K_{1,n}$ and the derived quantities is made clear by considering the co-ordinates of points on the ray in the object and image spaces at selected distances (measured along the ray) from the first and last surfaces. Let these distances measured in the positive direction from the surfaces be denoted by ρ and ρ', so that their co-ordinates satisfy

$$y=\eta_1+M_0\rho, \qquad y'=\eta_n+M_n\rho',$$
$$z=\zeta_1+N_0\rho, \qquad z'=\zeta_n+N_n\rho',$$

and therefore by (19)

$$y'+y\left(\frac{\rho'}{\mu_n}K_{1,n}-\frac{\partial K_{1,n}}{\partial K_n}\right)$$
$$=\mu_0M_0\left\{\frac{\rho}{\mu_0}\cdot\frac{\rho'}{\mu_n}K_{1,n}-\frac{\rho}{\mu_0}\frac{\partial K_{1,n}}{\partial K_n}+\frac{\rho'}{\mu_n}\frac{\partial K_{1,n}}{\partial K_1}-\frac{\partial^2 K_{1,n}}{\partial K_1\partial K_n}\right\}$$
$$=\frac{\mu_0M_0}{K_{1,n}}\left\{\left(\frac{\rho}{\mu_0}K_{1,n}+\frac{\partial K_{1,n}}{\partial K_1}\right)\left(\frac{\rho'}{\mu_n}K_{1,n}-\frac{\partial K_{1,n}}{\partial K_n}\right)+1\right\} \quad (29)$$

by (18). A similar relation holds if z and N are substituted for y and M. Now considerations of symmetry show that we should expect to find the image point in the same axial plane as the object point, or

$$\frac{y'}{y}=\frac{z'}{z}=G, \text{ say.}$$

If the image is plane and free from aberrations G is obviously its linear magnification compared with the object, and will be positive if the image is upright, negative if the image is inverted. Without assuming that aberrations are absent we may conveniently regard G as a magnification associated for a given object point with this particular ray.

§ (4) Conjugate Points.—Equation (29) shows that if by conjugate points we mean a pair of points on the ray, one in the object space and one in the image space, which lie in the same axial plane, their distances from the end surfaces are connected by the relation

$$\frac{\rho}{\mu_0}\cdot\frac{\rho'}{\mu_n}K_{1,n}-\frac{\rho}{\mu_0}\frac{\partial K_{1,n}}{\partial K_n}+\frac{\rho'}{\mu_n}\frac{\partial K_{1,n}}{\partial K_1}-\frac{\partial^2K_{1,n}}{\partial K_1\partial K_n}=0 \quad (30)$$

and that in terms of the magnification G,

$$\left.\begin{aligned}-\frac{\rho}{\mu_0}K_{1,n}&=\frac{\partial K_{1,n}}{\partial K_1}-\frac{1}{G}\\ \frac{\rho'}{\mu_n}K_{1,n}&=\frac{\partial K_{1,n}}{\partial K_n}-G\end{aligned}\right\}. \quad . \quad (31)$$

Particular instances of these results of great importance in the theory of optical instruments are that the image of the first surface of the instrument is within the image space at a distance

$$\mu_n\frac{\partial^2K_{1,n}}{\partial K_1\partial K_n}\bigg/\frac{\partial K_{1,n}}{\partial K_1}$$

from the last surface, and that for light travelling in the reverse direction the image

of the nth surface is in the first medium at a distance

$$\mu_0 \frac{\partial^2 K_{1,n}}{\partial K_1 \partial K_n} \Big/ \frac{\partial K_{1,n}}{\partial K_n}$$

from the first surface. Again, if the ray traced is a member of a pencil which is refracted telescopically, so that $K_{1,n}=0$, the magnification is independent of the distance of the object if this is finite, and is equal to $\partial K_{1,n}/\partial K_n$, or $1/(\partial K_{1,n}/\partial K_1)$. It is important, however, to note from (30) that the fact of a system being telescopic does not mean that the image of an object at a finite distance is not itself at a finite distance. On the contrary, the peculiarity of the telescope is that it is only objects which are themselves at infinity which have an image at an infinite distance from the instrument.

Conjugate points have so far been defined as the points in which a skew ray meets an axial plane. It is important to extend the definition in such a way as to have a definite meaning when the ray lies entirely in a plane. Equation (30) would serve as a basis for this purpose, for this equation remains determinate whether the ray lies in a plane or not. An alternative definition of importance which is consistent with this is obtained by eliminating $\partial K_{1,n}/\partial K_1$ and $\partial K_{1,n}/\partial K_n$ from (31) by means of (19) and (22). Thus

$$\left.\begin{aligned} \mu_0 M_0 - G\mu_n M_n &= GyK_{1,n} = y'K_{1,n} \\ \mu_0 N_0 - G\mu_n N_n &= GzK_{1,n} = z'K_{1,n} \end{aligned}\right\}, \qquad (32)$$

or eliminating G,

$$K_{1,n} = \frac{\mu_0 M_0}{y'} - \frac{\mu_n M_n}{y} = \frac{\mu_0 N_0}{z'} - \frac{\mu_n M_n}{z}. \qquad (33)$$

If a ray lies very nearly in an axial plane, say the plane $z=0$, all the N's and ζ's will be small quantities of the first order, and K will differ from the value for a ray actually in this plane by a second order quantity. It follows that the conjugate points for a ray in a plane are the points of intersection of the ray with the focal line which lies in the axial plane when the image is not free from aberration.

§ (5) Paraxial Rays and Collinear Images.—An example of the first importance of the application of the above formulae is for rays which meet each surface near its vertex and which always make small angles with the axis. In such cases, if second order quantities are neglected, every cos ϕ and cos ϕ' may be replaced by unity and every τ reduces to t, the axial thickness. It is convenient to denote the special value of the K in this instance by κ, so that

$$\kappa_m = (\mu_m - \mu_{m-1})R_m, \qquad (34)$$

and the four fundamental Gaussian constants of the system are derived by applying equations modelled on (16) or (20). The quantity $\kappa_{1,n}$ is called the power of the system, and its simplest interpretation is expressed in terms of the change of curvature an incident paraxial wave front undergoes on traversing the system.

Since $\kappa_{1,n}$ is invariable for all paraxial rays, whether skew or plane, it follows that the image of any plane object normal to the axis formed by such rays itself lies in a plane normal to the axis, and since all these rays enter and leave the system at points indistinguishable from points in the tangent planes to the first and last surfaces at their vertices, there will be no aberrations in the images formed by the rays, and the image will be identical with that which would result from collinear theories of imagery. This result, however, cannot in general hold for non-paraxial rays. For suppose there is no aberration over the whole plane image of a plane object when the magnification is G. By analogy with (31) the object and image planes for paraxial rays for magnification G′ will be displaced along the axis by

$$\mu_0 \frac{(1/G' - 1/G)}{\kappa_{1,n}} \text{ and } \mu_n \frac{G - G'}{\kappa_{1,n}}$$

from those for magnification G. It at once follows from (31) that if L_0 and L_n are the cosines of the inclination of the ray to the axis before and after refraction the magnification for the point in which this general ray meets the new object plane is G″, where

$$\left(\frac{1}{G''} - \frac{1}{G}\right)\frac{1}{K_{1,n}} = \left(\frac{1}{G'} - \frac{1}{G}\right)\frac{1}{L_0 \kappa_{1,n}},$$

and its conjugate point is not in the new image plane for paraxial rays, but at a perpendicular distance from it equal to

$$\frac{\mu_n (G - G')\{G(L_n \kappa_{1,n} - K_{1,n}) + G'(L_0 \kappa_{1,n} - K_{1,n})\}}{\kappa_{1,n}\{GK_{1,n} + G'(L_0 \kappa_{1,n} - K_{1,n})\}}.$$

The conjugate point will therefore only coincide with the collinear image point if

$$L_0 = L_n = \frac{K_{1,n}}{\kappa_{1,n}}, \qquad (35)$$

a relation which cannot hold for wide pencils of rays. It follows that collinear imagery in general is impossible, but that where an approach to it is of importance, as in photographic objectives, the most satisfactory results will be obtained by so arranging the stops that rays shall, as far as possible, lie after refraction parallel to their incident directions, or, in other words, the centres of the stops should lie close to the nodal points.

The impossibility of collinear imagery is seen more readily from (31) if the result given earlier but not yet proved is assumed, that $K_{1,n}$ must have a constant value for rays passing through any given point of the aberra-

tionless plane image. These equations at once show that the conjugate points for some other value of the magnification lie on the surfaces of spherical shells with the points in this plane as centres. The confused character of the image elsewhere is at once apparent on considering the aggregate of these surfaces. Evidently the more unequal the inclinations of the rays to the axis before and after refraction the more complete is the confusion out of the proper image plane for which the system is corrected, and hence follows the supreme importance of exact focussing with microscope objectives of large numerical aperture.

§ (6) PRINCIPLES OF IMAGE CORRECTION.—Equations (31) and (35) show that it is to be expected that $K_{1,n}$, while agreeing in sign with $\kappa_{1,n}$, should be numerically smaller in the case of all rays concerned with images in the outer parts of the field of view. On the other hand, for a single surface at which the deviation is δ equations (9) give

$$\left(\frac{K}{\kappa}\right)^2 = 1 + \frac{2\mu\mu'(1-\cos\delta)}{(\mu'-\mu)^2},$$

so that at every surface K tends to be greater than κ, which is its minimum value in refraction. This natural tendency to bad correction is accentuated where τ tends to be smaller than t, as in a simple converging lens. It can evidently be corrected in two ways: either there must be in the system some surfaces having powers differing in sign from that of the complete system, and at which the incidence of the marginal and oblique rays tends to be greater than in the case of surfaces of like power to that of the entire system, or, alternatively, the system may be constructed of separated lenses so arranged that the τ's for the marginal and oblique rays exceed the corresponding t's to an extent which more than compensates for the increased values of the K's due to the oblique incidence. As a rule where the principle of separation is employed the alternative principle is introduced as well. These two principles will provide a key to the design of many optical systems.

§ (7) THIN LENSES.—It will not be possible here to attempt any detailed discussion of the general problem of lens design, but on account of its importance in physical instruments some consideration will be given to the theory of "thin" lenses and the application of the results to lenses which it is possible to construct and which may be called "close" lenses, from the consideration that while their thicknesses are not negligible, yet they are kept as small as the attainment of the required aperture and power will allow, some margin being allowed in addition to secure the minimum substance necessary both for strength in use and for the rigidity in working on which the attainment of surfaces of good figure depends.

Consider, in the first place, a single lens of material of refractive index μ immersed in air. Suppose that the axial thickness is small, though not necessarily zero. The curvatures of the surfaces are R_1 and R_2. For convenience write ϖ for $1/\mu$. Then if ρ and ρ' are the distances from the two surfaces of the conjugate points for which $G=1$, equations (31) give

$$\rho K_{1,2} = \varpi\tau K_2, \quad \rho' K_{1,2} = -\varpi\tau K_1,$$

and similar formulae apply to paraxial rays with κ and t substituted for K and τ. If τ is not large, and the inclination of the ray to the axis is small, the distance of the ray from the axis will be approximately equal at the intersections with the surfaces of the lens and at the unit points as defined above. Denote the square of this distance by 2χ. Then these points will lie on spheres of curvatures σ and σ' through the axial unit points, where

$$\chi R_1 + \frac{\varpi\tau K_2}{K_{1,2}} = \frac{\varpi t\kappa_2}{\kappa_{1,2}} + \chi\sigma,$$

and
$$\chi R_2 - \frac{\varpi\tau K_1}{K_{1,2}} = -\frac{\varpi t\kappa_1}{\kappa_{1,2}} + \chi\sigma';$$

also
$$\chi R_1 + \tau = t + \chi R_2,$$

and therefore, neglecting to this approximation the differences between the K's and κ's as well as the squares of t and τ,

$$\sigma - \sigma' = (R_1 - R_2)(1-\varpi),$$
$$= \varpi\kappa_{1,2}, \quad . \quad . \quad . \quad . \quad (36)$$

if the thickness is neglected.

The unit surfaces therefore cannot both be planes in a thin lens. It should be noted that their relative curvature has the opposite sign to that of wave surfaces refracted paraxially, for the curvature of the refracted wave exceeds that of the incident wave by $\kappa_{1,2}$. Thus the statement sometimes made that a wave front, which at some instant coincides with some object surface will at some subsequent instant after refraction coincide with the corresponding image surface is incorrect. On the simplest general grounds, the idea that the time taken to form the image of, say, a plane surface is independent of the part of the image considered is evidently without foundation.

From (31) it is easily seen that if the suffixes 1 and 2 are used to denote two different pairs of conjugate points on a ray

$$\frac{\rho_1' - \rho_2'}{\mu_n} = \frac{\rho_1 - \rho_2}{\mu_0} G_1 G_2.$$

It follows that the magnification in the immediate neighbourhood of these surfaces

is unity in all directions, and that the curvature of any object near the first surface exceeds that of the image by the constant amount $\varpi\kappa_{1,2}$.

It remains to be seen how the exact value of the curvatures of these unit surfaces depends on the shape of the lens itself. From symmetry the curvatures will become equal and opposite in sign when the same is true of the refracting surfaces of the lens. On adding and simplifying, the equations giving σ and σ' lead to

$$\sigma+\sigma'=(R_1+R_2)(1+\varpi),$$

so that the increment of the curvature of each surface of the lens by a given amount increases the curvature of each unit surface by $(1+\varpi)$ times that amount. If the lens is thin and the curvature of each surface exceeds that of the equiconvex form by $S\kappa_{1,2}$ the curvatures of the unit surfaces are

$$\{\tfrac{1}{2}\varpi+S(1+\varpi)\}\kappa_{1,2} \text{ and } \{-\tfrac{1}{2}\varpi+S(1+\varpi)\}\kappa_{1,2}$$

respectively. These results can be readily derived from the consideration of a positive lens of small thickness the surfaces of which intersect when the aperture is small, if it is assumed that unit surfaces exist, that they are spherical and pass through the axial unit points and the circle in which the lens surfaces meet. It is evident from (31) and (32) that the shape of the unit surfaces, and hence the shape of the lens, affords means of altering the relation between M_0 and M_2 or N_0 and N_2 while maintaining a constant value for the paraxial power.

The conclusion that the unit surfaces are not planes prompts the inquiry whether a similar conclusion is true for images remote from the lens. The investigation is made most simply by considering a lens of zero thickness, and regarding the rays which are refracted at the common vertex of the surfaces. At the first surface

$$\cos\phi_1=1-\tfrac{1}{2}(M_0^2+N_0^2),$$

$$\cos\phi_1'=1-\frac{1}{2\mu^2}(M_0^2+N_0^2) \text{ approximately,}$$

and therefore

$$K_1=\kappa_1\left\{1+\frac{\varpi}{2}(M_0^2+N_0^2)+\ .\ .\ .\right\}.$$

Similarly at the second surface

$$K_2=\kappa_2\left\{1+\frac{\varpi}{2}(M_2^2+N_2^2)+\ .\ .\ .\right\},$$

but since refraction is at this point through a thin parallel film $M_2=M_0$, $N_2=N_0$. Thus for the lens as a whole

$$K_{1,2}=\kappa_{1,2}\left\{1+\frac{\varpi}{2}(M_0^2+N_0^2)+\ .\ .\ .\right\}, \quad (37)$$

and for magnification G the object surface in the neighbourhood of the axis is a sphere of curvature

$$\frac{(1+\varpi)\kappa_{1,2}}{1-1/G},$$

and the image surface a sphere of curvature

$$-\frac{(1+\varpi)\kappa_{1,2}}{1-G}.$$

The curvature of the object surface thus exceeds that of the image surface by the constant amount

$$(1+\varpi)\kappa_{1,2},$$

a result which it is easy to extend to the image of any spherical object since (30) takes the form

$$\frac{1}{\rho'}-\frac{1}{\rho}=K. \quad . \quad . \quad . \quad (38)$$

It will be borne in mind that in this discussion we have interpreted the "image" as the intersection of the ray with the focal line contained in an axial plane. It will be necessary later to consider the surfaces in which the focal line normal to an axial plane meets the ray, and the corresponding results will be found to differ from those derived above only in the numerical coefficients to be assigned to the characteristic terms.

It is desirable to extend these results to a system consisting of any number of thin lenses in contact with one another on the axis. For an oblique ray refracted through the system at the common vertex equations such as (37) and (38) apply to each individual lens, and since the K's are directly additive the above results may be at once applied to the complete system by substituting

$$(1+\varpi_{1,n})\kappa_{1,n}$$

for the special case with $n=2$, $\varpi_{1,n}$ being defined by

$$\varpi_{1,n}\kappa_{1,n}=\Sigma\varpi_m\kappa_{m,m+1}, \quad . \quad . \quad (39)$$

where the summation includes each thin lens of the system.

The results first established relating to the shape of the unit surfaces may now be extended. It is not permissible, without further examination, to infer that the ϖ appearing in the expressions for the curvatures of these surfaces in the case of a single lens may be generalised for a group of lenses according to the law expressed by (39). The axial thickness will be taken as zero, and the values of the τ's for a non-paraxial ray sufficiently small for the product of any two or more τ's to be negligible. The assumption that we are treating rays which traverse the system approximately at a constant distance from the axis gives

$$\tau_m=\chi(R_{m+1}-R_m),$$

and if the external media are of unit refractive index

$$\chi\sigma=\chi R_1+\rho,$$

$$\chi\sigma'=\chi R_n+\rho'.$$

To the approximation here required

$$\frac{\partial K_{1,n}}{\partial K_1}=1-\Sigma\varpi_m\tau_m\kappa_{m+1,n},$$

$$\frac{\partial K_{1,n}}{\partial K_n}=1-\Sigma\varpi_m\tau_m\kappa_{1,m},$$

$$K_{1,n}=\kappa_{1,n},$$

and therefore

$$\left.\begin{aligned}\sigma&=R_1+\frac{1}{\kappa_{1,n}}\Sigma\varpi_m(R_{m+1}-R_m)\kappa_{m+1,n}\\ \sigma'&=R_n-\frac{1}{\kappa_{1,n}}\Sigma\varpi_m(R_{m+1}-R_m)\kappa_{1,m}\end{aligned}\right\}, \quad (40)$$

where the summations extend over the internal media, *i.e.* from $m=1$ to $m=n-1$. Considering first the curvature difference

$$\begin{aligned}\sigma-\sigma'&=R_1-R_n+\Sigma\varpi_m(R_{m+1}-R_m)\\ &=\sum_1^n R_m(\varpi_{m-1}-\varpi_m)\\ &=\sum_1^n \kappa_m\varpi_{m-1}\varpi_m\\ &=\varpi_{1,n}\kappa_{1,n}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (41)\end{aligned}$$

by (39) since we are regarding, as we may do without any loss of generality, that every alternate medium is that in which the whole system of thin lenses is immersed, and thus has its ϖ equal to unity.

By addition (40) gives

$$\sigma+\sigma'=R_1+R_n$$
$$-\frac{1}{\kappa_{1,n}}\Sigma\varpi_m(R_{m+1}-R_m)(\kappa_{1,m}-\kappa_{m+1,n}).$$

It is convenient to group together the curvatures in pairs, taking together each odd m with the next higher even integer. Each ϖ of even suffix is unity. The equation then reduces immediately to

$$(\sigma+\sigma')\kappa_{1,n}=\Sigma(R_m+R_{m+1})\kappa_{m,m+1}(1+\varpi_m)$$
$$+\Sigma\varpi_m\kappa_{m,m+1}(\kappa_{1,m-1}-\kappa_{m+2,n}),$$

where the summation is taken for all the individual lenses, m having all odd values from 1 to $n-1$ inclusive. Since σ and σ' have the dimensions of a curvature, we may write

$$\sigma+\sigma'=\beta\kappa_{1,n}, \quad . \quad . \quad . \quad . \quad (42)$$

where β is a pure number, and if we take S_m as the deformation of the typical lens from its symmetrical form we find, with a slight alteration in the notation,

$$\beta\kappa_{1,n}^2=2\Sigma S_m(1+\varpi_m)\kappa_m^2$$
$$+\Sigma\varpi_m\kappa_m(\kappa_{1,m}-\kappa_{m,n}), \quad (43)$$

where now κ_m is the power of a complete component lens, and not of a single surface.

If the curvature of every surface in the system is increased by the same amount $S\kappa_{1,n}$, it is evident that β is increased by

$$2S(1+\varpi_{1,n}), \quad . \quad . \quad . \quad (44)$$

a result in complete accordance with that found for a single lens. It is not difficult to establish the generalised results (41), (42) from those for the single lens in their general form by so bending the system as a whole that the corresponding unit surfaces of the components have the same curvature while the required difference of curvature in the boundaries of the air lens is maintained.

It would be quite possible to take as the standard form for reference in this compound system that configuration in which the curvatures of the unit surfaces are equal and opposite, but it is found more convenient to adopt the form determined by the condition

$$\beta\kappa_{1,n}^2=\Sigma S_m\kappa_m^2, \quad . \quad . \quad . \quad (45)$$

for reasons which will shortly be apparent. In the lens combinations found in actual instruments the values of β determined by this condition are invariably small.

The fact that the curvatures of the unit surfaces have been found in a form which only involves the powers for paraxial rays shows that the results obtained apply for all rays, whether skew or otherwise, which satisfy the assumptions made in this investigation. In particular no distinction need be drawn between the two focal lines, the results thus differing markedly from those which hold for the images of objects at some distance from the lens. The unit surfaces in the neighbourhood of the axis of a thin lens are entirely free from astigmatism, and indeed it is this fact which makes the presence of astigmatism elsewhere unavoidable. It appears necessary for astigmatism to be present throughout the field of such systems as are of practical importance, with the possible exception of limited regions, and to obtain freedom from astigmatism where it is required, as in photographic lenses, it is necessary to introduce it in the neighbourhood of the lenses themselves. It is for this reason that in systems where the correction of astigmatism is desired a number of lenses well separated from one another are employed. Such systems tend to be the reverse of compact, and much skill is called for in combining to as great an extent as is possible the contradictory requirements of a compact lens and a field corrected for astigmatism.

§ (8) THIN LENSES—DETAILED CONSIDERATION.—Having now located the unit surfaces for any thin system it is next necessary to consider how $K_{1,n}$ varies not merely with the obliquity of the ray but also with the distance of its point of incidence from the axis. Approximate values of the τ's are known, and it has been shown that the difference between the K for any surface and the corresponding κ depends upon $1-\cos\delta$, where δ is the deviation at this surface. It is important that all the new terms should be expressed in terms of the external independent variables, and equations (19) and (22) will first be transformed to express the direction and position

of the ray at an intermediate stage of its passage through the system in terms of its initial and final co-ordinates. Let η and ζ without any suffix indicate the co-ordinates derived from (31) by putting $G=1$. Then η and ζ may relate at will to either the object or the image space. Equations (32) give

$$\left.\begin{aligned}\eta K_{1,n}&=\mu_0 M_0-\mu_n M_n\\ \zeta K_{1,n}&=\mu_0 N_0-\mu_n N_n\end{aligned}\right\} \quad . \quad . \quad . \quad (46)$$

Equations of the form of (19) and (22) provide independent sets of equations for any of the intermediate quantities, and on simplifying by using well-known relations there result the relations

$$\left.\begin{aligned}\mu_m M_m K_{1,n}&=\mu_0 M_0 K_{m+1,n}+\mu_n M_n K_{1,m}\\ \eta_m K_{1,n}&=\mu_0 M_0\frac{\partial K_{m,n}}{\partial K_m}-\mu_n M_n\frac{\partial K_{1,m}}{\partial K_m}\end{aligned}\right\}, \quad (47)$$

with similar equations for N_m and ζ_m.

For the special case of a thin system it is convenient to rearrange these equations, substituting κ's for K's in the form

$$2\mu_m M_m=\mu_0 M_0+\mu_n M_n-\eta(\kappa_{1,m}-\kappa_{m+1,n}),$$

so that

$$\begin{aligned}2(M_m-M_{m-1})=\frac{\mu_m-\mu_{m-1}}{\mu_m\mu_{m-1}}[&-(\mu_0 M_0+\mu_n M_n)\\ &+\eta\{\kappa_{1,m}-\kappa_{m,n}-R_m(\mu_m+\mu_{m-1})\}]\end{aligned}$$

and therefore

$$\begin{aligned}K_m=\kappa_m+\frac{\kappa_m}{8\mu_m\mu_{m-1}}[&(\mu_0 M_0+\mu_n M_n)^2+(\mu_0 N_0+\mu_n N_n)^2\\ &-2\{(\mu_0 M_0+\mu_n M_n)\eta+(\mu_0 N_0+\mu_n N_n)\zeta\}\\ &\times\{\kappa_{1,m}-\kappa_{m,n}-R_m(\mu_m+\mu_{m-1})\}\\ &+(\eta^2+\zeta^2)\{\kappa_{1,m}-\kappa_{m,n}-R_m(\mu_m+\mu_{m-1})\}^2].\end{aligned}$$

Now introduce the conditions that the external media are of unit refractive index, and suppose also that the same is true of the media ${}_{m-1}$ and ${}_{m+1}$. Then after slight simplification

$$\begin{aligned}K_m+K_{m+1}=\kappa_{m,m+1}&\\ &+\tfrac{1}{8}\{(M_0+M_n)^2+(N_0+N_n)^2\}\varpi_m\kappa_{m,m+1}\\ &+\tfrac{1}{4}\{(M_0+M_n)\eta+(N_0+N_n)\zeta\}\{(R_m+R_{m+1})(1+\varpi_m)\\ &-(\kappa_{1,m-1}-\kappa_{m+2,n})\varpi_m\}\kappa_{m,m+1}\\ &+\tfrac{1}{8}(\eta^2+\zeta^2)\{4R_m^2\kappa_m+4R_{m+1}^2\kappa_{m+1}\\ &-2(R_m+R_{m+1})(\kappa_{1,m-1}-\kappa_{m+2,n})(1+\varpi_m)\kappa_{m,m+1}\\ &+2(R_m-R_{m+1})(1+\varpi_m)\kappa_m\kappa_{m+1}\\ &+(\kappa_{1,m-1}-\kappa_{m+2,n})^2\varpi_m\kappa_{m,m+1}\\ &+\varpi_m(\kappa_m^2+\kappa_{m+1}^2)\kappa_{m,m+1}\}.\end{aligned}$$

The simplification of the last term will not yield the correct coefficient for $(\eta^2+\zeta^2)$ on summing up for the component lenses, for $K_{1,n}$ is less than ΣK_m by the amount

$$\Sigma\frac{\tau_m}{\mu_m}\kappa_{1,m}\kappa_{m+1,n},$$

and the terms in this sum which involve R_m and R_{m+1} are

$$\begin{aligned}&\tfrac{1}{2}(\eta^2+\zeta^2)[R_m\kappa_{1,m-1}\kappa_{m,n}\\ &\quad+(R_{m+1}-R_m)\varpi_m\kappa_{1,m}\kappa_{m+1,n}-R_{m+1}\kappa_{1,m+1}\kappa_{m+2,n}]\\ &\quad=\tfrac{1}{8}(\eta^2+\zeta^2)[(R_m-R_{m+1})(1-\varpi_m)\kappa_{1,n}^2\\ &\qquad-R_m(\kappa_{1,m-1}-\kappa_{m,n})^2\\ &\qquad-(R_{m+1}-R_m)\varpi_m(\kappa_{1,m}-\kappa_{m+1,n})^2\\ &\qquad+R_{m+1}(\kappa_{1,m+1}-\kappa_{m+2,n})^2]\\ &=\tfrac{1}{8}(\eta^2+\zeta^2)[\varpi_m\{\kappa_{1,n}^2-(\kappa_{1,m-1}-\kappa_{m+2,n})^2\}\\ &\qquad+2(R_m+R_{m+1})(1+\varpi_m)(\kappa_{1,m-1}-\kappa_{m+2,n})\\ &\qquad-(R_m-R_{m+1})\kappa_{m,m+1}\\ &\qquad+(1-\varpi_m)(R_m+R_{m+1})^2]\kappa_{m,m+1}.\end{aligned}$$

On taking these terms into account, it is found that

$$\begin{aligned}K_{1,n}=\kappa_{1,n}&+\tfrac{1}{8}\{(M_0+M_n)^2+(N_0+N_n)^2\}\varpi_{1,n}\kappa_{1,n},\\ &+\tfrac{1}{4}\{(M_0+M_n)\eta+(N_0+N_n)\zeta\}\beta\kappa_{1,n}^2,\\ &+\tfrac{1}{8}(\eta^2+\zeta^2)(\gamma-\varpi_{1,n}-1)\kappa_{1,n}^3, \quad . \quad . \quad (48)\end{aligned}$$

where, since

$$\kappa_{1,n}^3=\Sigma\kappa_m^3+3\Sigma(\kappa_{1,m}-\kappa_{m,n})^2\kappa_m,$$

$$\begin{aligned}\gamma\kappa_{1,n}^3=\Sigma\Bigl\{\Bigl(\frac{\kappa_m}{1-\varpi_m}\Bigr)^2&+4(1+2\varpi_m)S_m^2\kappa_m^2\\ &-8(1+\varpi_m)S_m\kappa_m(\kappa_{1,m}-\kappa_{m,n})\\ &+(3+2\varpi_m)(\kappa_{1,m}-\kappa_{m,n})^2\Bigr\}\kappa_m, \quad (49)\end{aligned}$$

and each κ now relates to a separate thin lens.

When the curvature of every surface in the system is increased by the same amount $S\kappa_{1,n}$, $\varpi_{1,n}$ remains constant, and the change in β has already been considered. The increase in $\gamma\kappa_{1,n}^3$ is, by (49),

$$\begin{aligned}4S^2\kappa_{1,n}^2&\Sigma(1+2\varpi_m)\kappa_m\\ &+8S\kappa_{1,n}\Sigma\{(1+2\varpi_m)S_m\kappa_m^2\\ &-(1+\varpi_m)(\kappa_{1,m}-\kappa_{m,n})\kappa_m\}\end{aligned}$$

or

$$4(1+2\varpi_{1,n})S^2\kappa_{1,n}^3+8(\beta\kappa_{1,n}^2-\Sigma S_m\kappa_m^2)S\kappa_{1,n},$$

and the meaning of the selection of condition (45) for the determination of the standard form is that we are choosing the configuration for which γ is a minimum. Any general curvature change from this form alters β according to a linear and γ according to a parabolic law.

It is easy to discover a meaning for the new quantity γ. Consider a beam of light arising from a point on the axis placed to give an inverted image of a small transverse object equal to the object itself for paraxial rays. To the approximation here required the refracted ray lies in the direction determined by $M_0+M_n=N_0+N_n=0$. Let A (*Fig.* 4) be the object point, B its paraxial image, and B′ the point in which a non-paraxial ray meets the axis. Let this ray meet the unit surfaces

whose vertices are at C, C′ in the points P, P′ and the spheres through C and C′ with centres

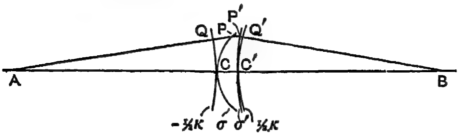


FIG. 4.

A and B in Q and Q′. Then Q, P, P′, Q′ lie on spheres of curvatures

$$-\tfrac{1}{2}\kappa_{1,n},\ \sigma, \sigma', \tfrac{1}{2}\kappa_{1,n},$$

so that

$$QP = \tfrac{1}{2}(\tfrac{1}{2}\kappa_{1,n} + \sigma)(\eta^2 + \zeta^2),$$

$$P'Q' = \tfrac{1}{2}(\tfrac{1}{2}\kappa_{1,n} - \sigma')(\eta^2 + \zeta^2),$$

and

$$\frac{1}{AP} + \frac{1}{P'B} = \frac{1}{(2/\kappa_{1,n}) + QP} + \frac{1}{(2/\kappa_{1,n}) + P'Q'}$$

$$= \kappa_{1,n} - \tfrac{1}{8}(\kappa_{1,n} + \sigma - \sigma')\kappa_{1,n}^2(\eta^2 + \zeta^2)$$

$$= \kappa_{1,n} - \tfrac{1}{8}(1 + \varpi_{1,n})\kappa_{1,n}^3(\eta^2 + \zeta^2) \text{ by (41);}$$

or, on comparison with (48),

$$\tfrac{1}{4}B'B\kappa_{1,n}^2 = \frac{1}{P'B'} - \frac{1}{P'B} = \tfrac{1}{8}(\eta^2 + \zeta^2)\gamma\kappa_{1,n}^3,$$

showing that γ is a numerical measure of the spherical aberration for the magnification -1. When γ is positive the axial intersection point B′ is displaced towards the lens from the paraxial position B, and the aberration is conventionally described as positive or undercorrected.

If $\gamma = 0$ the axial aberration is removed for this image point, and K falls as the aperture increases. The apparent contradiction between this result and that given in § (3), to the effect that K should be constant for all rays from a given object point, is due to the impossibility with a thin lens system of achieving an aberrationless plane image of a plane object.

Since the variation of the power and the shapes of the unit surfaces have been expressed in terms of three quantities—ϖ, β, γ—these suffice to determine completely to this order of accuracy the properties of the system for all object positions. The calculation of any thin system therefore resolves itself into the problem of securing for these quantities arbitrarily assigned values, and it is important to know what values must be adopted to attain the effects required. A slight generalisation of the investigation into the meaning of γ will suffice for this purpose. If the object and image surfaces for magnification G are considered, and it is assumed that these have uniform curvatures s and s', the length of the ray between the unit surface at $\frac{1}{2}(\eta^2 + \zeta^2)\sigma$, η, ζ and the object at $(1/G - 1)/\kappa + \frac{1}{2}(y'^2 + z'^2)s/G^2$, y'/G, z'/G being ρ and the corresponding image length between $\frac{1}{2}(\eta^2 + \zeta^2)\sigma'$, η, ζ and $(1 - G)/\kappa + \frac{1}{2}(y'^2 + z'^2)s'$, y', z' being ρ', it is readily seen that

$$\frac{1}{\rho'} - \frac{1}{\rho} = \kappa$$

$$-\frac{\kappa^2}{2(1-G)^2}[\{\sigma G^2 - \sigma' + \kappa(1 + G + G^2)\}(\eta^2 + \zeta^2)$$

$$-2\kappa(1 + G)(\eta y' + \zeta z') + (\kappa - s + s')(y'^2 + z'^2)].$$

Also from (32) and (46)

$$M_0 + M_n = \frac{2y' - \eta(1 + G)}{1 - G}K,$$

with a similar equation for the N's, and the substitution of these values in (48), on replacing the K's by their approximate κ values, yields

$$K = \kappa + \frac{\kappa^3}{8}\Big[\frac{4\varpi}{(1-G)^2}(y'^2 + z'^2)$$

$$+ \frac{4}{1-G}\left(\beta - \varpi\frac{1+G}{1-G}\right)(\eta y' + \zeta z')$$

$$+ \left\{\gamma - \varpi - 1 - 2\beta\frac{1+G}{1-G} + \varpi\left(\frac{1+G}{1-G}\right)^2\right\}(\eta^2 + \zeta^2)\Big].$$

The relation (38) then gives, as the necessary conditions for an extended object-image relation between the spherical object and image the spherical aberration condition,

$$\gamma - 4\beta\frac{1+G}{1-G} + (3 + 2\varpi)\left(\frac{1+G}{1-G}\right)^2 = 0, \quad (50)$$

a second condition, which will subsequently be identified with the sine condition,

$$\beta - (2 + \varpi)\frac{1+G}{1-G} = 0 \quad . \quad . \quad (51)$$

and the curvature relation

$$s - s' - (1 + \varpi)\kappa = 0. \quad . \quad . \quad (52)$$

The discrepancy between (41) and (52) should be particularly noted. The explanation is evident if it is noted that, whatever G may be, (52) holds if $\eta = \zeta = 0$, and that it is assumed in obtaining this relation that the rays involved are those which are incapable of predicting what the curvature of the unit surfaces will be.

If the value of β determined by (51) is substituted in (50) it is seen that

$$\gamma - (5 + 2\varpi)\left(\frac{1+G}{1-G}\right)^2 = 0. \quad . \quad . \quad (53)$$

Introduce now the special convention as regards β and γ that the addition of the suffix $_0$ signifies that the coefficient relates to the system in its zero configuration. Then if the curvature addition from this form is $S\kappa_{1,n}$

$$\gamma = \gamma_0 + 4(1 + 2\varpi)S^2,$$

$$\beta = \beta_0 + 2(1 + \varpi)S.$$

If in this form β and γ have the values required by (51) and (53) the elimination of S gives

$$\left(\frac{1+G}{1-G}\right)^2+2\frac{1+G}{1-G}(1+2\varpi)(2+\varpi)\beta_0 -(1+2\varpi)\beta_0^2-(1+\varpi)^2\gamma_0=0,$$

showing that bending the system as a whole will in any given case only yield two object positions, other than contact with the lens, for which the satisfaction of the sine condition is consistent with the removal of spherical aberration.

These two solutions are both imaginary unless

$$(1+2\varpi)(5+2\varpi)\beta_0^2+\gamma_0$$

is positive, and this condition cannot be satisfied with most combinations of two glasses to form a cemented achromatic doublet. Although both these conditions should be satisfied in such a system as a telescope objective, it frequently happens that the sine condition is deliberately ignored because the use of other glasses simplifies the construction from the manufacturing standpoint, and the faults present are not very conspicuous when the field of view is small.

§ (9) The Cemented Doublet.—On account of the great importance of lenses of the cemented doublet class it is desirable to consider them in some detail.

The powers of the component lenses are determined by the assigned focal length or power of the combination and the colour conditions which are to be satisfied. It is therefore assumed that these powers are known.

From equations (43) and (45) it is evident that the common curvature addition required to attain the standard form is

$$-\frac{1}{(1+2\varpi)\kappa}\sum_1^n\{(1+2\varpi_m)S_m\kappa_m -\varpi_m(\kappa_{1,m}-\kappa_{m,n})\}\kappa_m,$$

where in the absence of a suffix $_{1,n}$ is to be understood. It at once follows that

$$\beta_0(1+2\varpi)\kappa^2=\sum_1^n(\varpi-\varpi_m)(2S_m\kappa_m-\kappa_{1,m}+\kappa_{m,n})\kappa_m,$$

and since the range of indices of available glasses is very restricted the first factor on the right is normally small, and β_0 is a small quantity. The small range in the possible values of ϖ and β_0 involves the conclusion that the main problem concerns the attainment of a suitable value for γ_0, either by the initial selection of suitable varieties of glass or by an arbitrary division of the system. If Q_m is written for $2S_m\kappa_m-\kappa_{1,m}+\kappa_{m,n}$, k_m for $(1+2\varpi_m)\kappa_m$, and j_m for $\kappa_m/(1-\varpi_m)$, and the double suffix notation is used to denote summation, the equations for β_0 and γ_0 may be written

$$2\beta_0\kappa=\frac{1}{\kappa}\Sigma Q_m\kappa_m-\frac{1}{k}\Sigma Q_m k_m, \quad . \quad (54)$$

$$\gamma_0\kappa^3=\Sigma\{j_m^2\kappa_m+Q_m^2k_m-2Q_m(\kappa_{1,m}-\kappa_{m,n})\kappa_m\} -\frac{1}{k}\{\Sigma Q_m k_m\}^2, \quad . \quad (55)$$

where the suffix $_{1,n}$ is to be understood when none is given. It is evident by inspection that the equation for β_0 is unaffected by making any common addition to all the Q's, and the same becomes obvious for γ_0 by writing the last equation in the form

$$\gamma_0\kappa^3=\Sigma\{j_m^2\kappa_m+\frac{1}{k}k_l k_m(Q_l-Q_m)^2 +2(Q_m-Q_{m+1})\kappa_{1,m}\kappa_{m+1,n}\}.$$

Now assume for Q_m the form

$$-Q_m=P_{1,m-1}-P_{m,n-1}+j_{1,m}-j_{m,n},$$

where $P_1, P_2, \ldots P_{n-1}$ are $(n-1)$ unknown quantities and the double suffix carries its general additive meaning. Substituting for the Q's and j's in terms of the S's and κ's gives

$$S_m\kappa_m-\frac{\kappa_m}{2(\mu_m-1)}=P_m+S_{m+1}\kappa_{m+1}+\frac{\kappa_{m+1}}{2(\mu_{m+1}-1)}$$

and

$$S_1\kappa_1+\frac{\kappa_1}{2(\mu_1-1)}+S_n\kappa_n-\frac{\kappa_n}{2(\mu_n-1)}=0,$$

showing that the curvatures of the extreme surfaces of the system as given by the Q's are equal and opposite, and that P_m is the curvature difference between the consecutive surfaces belonging to lenses $_m$ and $_{m+1}$. For a cemented system the P's will all be zero, and γ_0 may be calculated from

$$\gamma_0\kappa^3=\Sigma\left\{j_m^2\kappa_m+\frac{1}{k}k_l k_m(j_{l,m-1}+j_{l+1,m})^2 +2(j_m+j_{m+1})\kappa_{1,m}\kappa_{m+1,n}\right\}.$$

When the system consists of only two members convenient expressions for numerical use are

$$\beta_0\kappa^2k=-(\varpi_1-\varpi_2)\kappa_1\kappa_2 j, \quad . \quad . \quad (56)$$

which illustrates the general rule that unlike ϖ and γ_0, which are unaltered, β_0 changes sign when the system is reversed, and

$$\gamma_0\kappa^2=j_1^2\kappa_1+j_2^2\kappa_2+2j\kappa_1\kappa_2+k_1k_2j^2/k. \quad (57)$$

In an achromatic objective one component is positive and one negative, while κ, j, k agree in sign. Thus only one of the terms gives a positive contribution to γ_0, which has a strong tendency to be negative. In this respect a doublet is in marked contrast with a single lens, which necessarily has a positive γ_0. While this difference enables the combination to be corrected for spherical aberration, it usually prevents the simultaneous

correction of spherical aberration and the sine condition. From the condition obtained earlier it is evident that for the simultaneous correction γ_0 when negative must be numerically small. It is seen on considering the results for a cemented doublet and a single lens that the fault lies in too great a difference of refractive index between the two glasses employed.

To obtain a general idea of the kind of variation of γ_0 implied by the above expression, assume that the refractive indices of the two glasses are given and that the power of the cemented combination is unity. Consider the change in γ_0 as the power of one component falls from a large positive to a large negative value, and the other consequently rises from a large negative to a large positive value. Let component 1 have the lower refractive index, that is $\varpi_1 > \varpi_2$. At the extremes the sign of γ_0 is given by the highest power of $\kappa_1 - \kappa_2$ and is obtained by writing $\frac{1}{2}(\kappa_1 - \kappa_2)$ for κ_1, and $-\frac{1}{2}(\kappa_1 - \kappa_2)$ for κ_2. The result is the value

$$-\frac{(\kappa_1-\kappa_2)^3(\varpi_1-\varpi_2)(1+8\varpi_1\varpi_2)}{16(1-\varpi_1)^2(1-\varpi_2)^2},$$

showing that with powerful lenses of opposite signs γ_0 tends to assume a positive value when the component with the higher index is of the same sign as the combination, and a negative value when the component with the same sign as the combination has the lower index. These extreme values illustrate the general tendency of the change in the series, and agree with the well-known construction for telescope objectives in which the combination and the lens of lower refractive index are positive or converging, and the system is corrected for spherical aberration by bending to an extent dependent on the magnitude of the negative γ_0. It must not be thought, however, that the variation in γ_0 is throughout uniform in direction. For taking the series in the direction from γ_0 negative to γ_0 positive we reach $\kappa_2=0$, $\kappa_1=1$ before $\kappa_1=0$, $\kappa_2=1$. In these cases γ_0 reduces to the single-lens formula $\gamma_0=(1-\varpi)^{-2}$, and since $0<\varpi<1$ and $\varpi_1>\varpi_2$, the value of γ_0 in the earlier case exceeds that in the later. There must therefore exist within the range positions where γ_0 becomes stationary. In the neighbourhood of $\kappa_2=0$ equation (57) gives

$$\gamma_0=\frac{1}{(1-\varpi_1)^2}\left[1-\frac{2\kappa_2}{\kappa}(\varpi_1-\varpi_2)\ .\ .\ .\right],$$

showing that stationary points exist outside the region in which both components have the same sign. The stationary point near this region for which γ_0 is a minimum would be of great importance if γ_0 reached values in the neighbourhood of zero, but this is found not to be the case. A typical curve exhibiting the variation of γ_0 in the region of practical interest is shown for $\varpi_1=\frac{2}{3}$, $\varpi_2=\frac{5}{8}$, corresponding to $\mu_1=1{\cdot}5$, $\mu_2=1{\cdot}6$ (*Fig.* 5). The values of

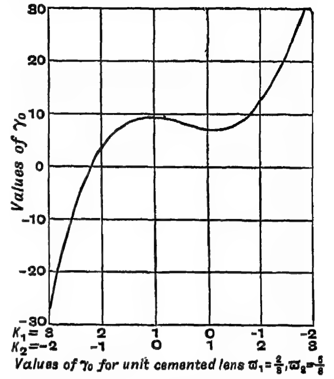


FIG. 5.

γ_0 at the stationary points do not differ much from those when one component falls to zero power.

If a number of curves of this kind are constructed, with one of the glasses of fixed index while the other is variable, it becomes possible to exhibit the variation of γ_0 for cemented doublets, when one independent variable is the refractive index that is not fixed, or a convenient function of this index, and the other determines the relative magnitude of the powers of the two component lenses, by drawing lines on a chart, in which those variables are represented by distances measured in perpendicular directions, through the points which have common values of γ_0. The contour map in which one of the variables is $(\kappa_1-\kappa_2)/(\kappa_1+\kappa_2)$ is most instructive, but for practical use in detailed work it is more convenient to substitute $\log(-\kappa_1/\kappa_2)$. The two maps are of course quite dissimilar, and it is not easy with the latter variable to trace the effect of a continuous variation of one of the component powers from positive through zero to negative values. Its special advantage is that it enables an important group of problems to be solved rapidly by showing which pairs of glasses are suitable for the purpose in hand. It is supposed that the colour conditions are given in a form which fixes the ratio of the component powers in terms of the "constringence" of the glasses, *i.e.* the coefficient ν. If, then, two charts are prepared, one exhibiting the contour lines for γ_0, and one representing by points the various glasses known to be available, both charts having the refractive index represented on the same scale, and the other variable being in the first case $\log(-\kappa_1/\kappa_2)$, and in the other $\log \nu_1/\nu_2$ or other appropriate representation of the colour

condition required, the best pair of glasses can be selected by superposing the two charts and moving them in slide-rule fashion over one another, provided the upper one is drawn on transparent material. The accompanying figures show such a pair of charts for finding

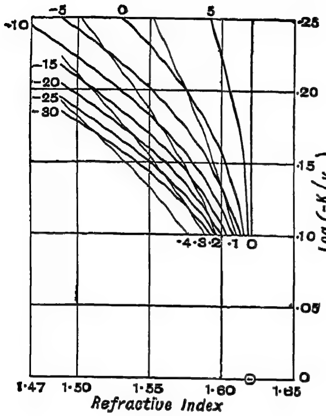


FIG. 6.

glasses fully corrected for colour. The fixed refractive index is 1·62, but the same charts may be used for a considerable range of values for this index. It will be noted that to secure

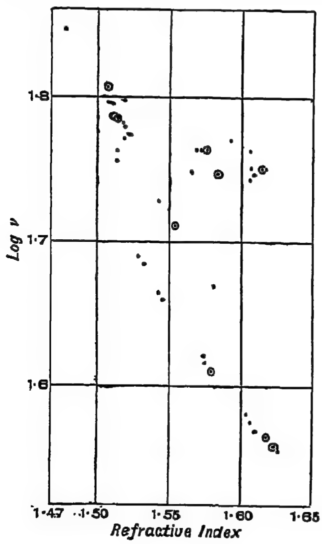


FIG. 7.

simultaneous correction for colour, spherical aberration, and coma in a thin cemented objective the precise dense flint selected is of little importance, but that in the crowns the index of the ordinary silicate glasses is too small while that of the dense barium crowns is too great. Of intermediate glasses the medium barium crowns are suitable, as are also the phosphate crowns introduced in 1886 at Jena, but since withdrawn on account of their instability, and also the fluor crowns introduced in 1913.

§ (10) EXTENSION OF THE CEMENTED DOUBLETS.—When suitable glasses are not available to meet the conditions that must be satisfied, two courses are open, both of which have as their object the securing of an additional degree of freedom. One of these consists in removing the restriction that the inner curvatures of the two components are to be complementary, so that the lenses may be cemented together; the other retains the condition that the inner surfaces are to be cemented, and secures the required freedom by dividing the total power to be contributed by one of the two glasses into two arbitrary parts, the final lens thus consisting of the lens of the other glass cemented between two lenses of the glass that has been divided.

Considering first the case in which the requirement that there shall be no air-gap is abandoned, the value $\bar{\gamma}_0$ for γ_0 when an air-gap P is present is given by

$$(\bar{\gamma}_0-\gamma_0)\kappa^3 k_1 k_2 k=(2\mathrm{P}k_1k_2+\kappa_1\kappa_2 k+k_1k_2 j)^2 -(\kappa_1\kappa_2 k+k_1k_2 j)^2.$$

In the cases of practical importance k has the same sign as κ, and k_1 and k_2 differ in sign. The introduction of the air-gap thus gives a limited opportunity of increasing γ_0, and offers a theoretically unlimited range of smaller values of γ_0. It is thus obvious that when an air-gap may be introduced any one aberration condition can always be satisfied, a result which does not apply to a cemented objective of two components, and as a rule two conditions can be complied with. The problem reduces in all cases to the solution of a quadratic equation, since one condition gives a linear and the other a quadratic relation between the amounts to which the two components must be bent, or, what is equivalent, between the bending of the complete lens and the air-gap curvature.

When one of the lenses is to be divided the cemented system that is to be determined may either be regarded as two achromatic cemented lenses cemented together with their like glasses in contact,[1] or treated directly as a triple lens without air-gaps. If κ_1, κ_2, κ; j_1, j_2, j; k_1, k_2, k have exactly the same values as they bore for the doublet lens, the elements for the triple lens may be represented by $\kappa_1(1-\theta)$, κ_2, $\kappa_1\theta$; $j_1(1-\theta)$, j_2, $j_1\theta$; $k_1(1-\theta)$, k_2, $k_1\theta$, and the values of the Q's will be $j-j_1(1-\theta)$,

[1] See *Proc. Phys. Soc. Lond.* xxvii. 405.

$-j_1(1-2\theta)$, $-j+j_1\theta$. The substitution of these values in (54) and (55) leads at once to

$$\bar{\beta}_0=\beta_0(1-2\theta) \quad . \quad . \quad . \quad (58)$$

and $$\bar{\gamma}_0-A'=(\gamma_0-A')(1-2\theta)^2, \quad . \quad (59)$$

where

$$A'\kappa^2=(\kappa\kappa_1+jk_1)j_2+j_1^2\kappa_1+j_2^2\kappa_2+\kappa_1\kappa_2 j$$

$$=j^2\kappa-j_1j_2(\varpi_1-\varpi_2)(j+j_2)\varpi_1, \quad . \quad (60)$$

and the coefficients with a bar above relate to the triple lens, while those without refer to the cemented doublet. It will be noted that the last three terms in the first expression for A′ have been found in evaluating γ_0. The meaning of A′ becomes evident at once on putting $\theta=\frac{1}{2}$, which gives $\bar{\gamma}_0=A'$. Since the lens is then quite symmetrical in construction, and $\bar{\gamma}_0=0$ is the condition for freedom from spherical aberration at magnification -1, it follows that $A'=0$ is the condition that a parallel beam incident on the second component of the doublet with the external surface flat should be brought to a focus without aberration.

The corresponding result when the second component of the doublet is divided to form two external lenses may be written down by analogy. If the three components are $\kappa_2\theta$, κ_1, $\kappa_2(1-\theta)$ equation (58) is unaltered, and in (59) A′ must be replaced by A where

$$A\kappa^2=(\kappa\kappa_2+jk_2)j_1+j_1^2\kappa_1+j_2^2\kappa_2+\kappa_1\kappa_2 j$$

$$=j^2\kappa+j_1j_2(\varpi_1-\varpi_2)(j+j_1)\varpi_2,$$

and represents the spherical aberration for an object at infinity when the light is incident on the cemented doublet 1, 2 with the first surface plane.

It is evident that a triple lens may have any value of γ_0 on the same side of A′ or A, as the case may be, as the γ_0 of the cemented doublet. As a rule it will be possible to satisfy two conditions, and four solutions will in general be found for any given problem, two of which have external lenses of one glass, and two of the other.

These results may readily be generalised to yield simple equations for the solution of any problem with two given glasses when the lens is to consist of any number of thin components cemented together, the two glasses occurring alternately.[1] In all cases a solution may be obtained by solving a quadratic equation.

§ (11) A Telescope Objective free from Spherical Aberration and Coma.—As examples of the application of the foregoing expressions let it be required to find the curvatures for lenses free from spherical aberration, and where possible from coma also, for an object at infinity, with

$$\mu_1=1{\cdot}5, \quad \mu_2=1{\cdot}6, \quad \kappa_1=3, \quad \kappa_2=-2.$$

Then

$$\varpi_1=\tfrac{2}{3}, \quad \varpi_2=\tfrac{5}{8}, \quad j_1=9, \quad j_2=-\tfrac{16}{3}, \quad k_1=7, \quad k_2=-\tfrac{9}{2}.$$

For a cemented doublet

$$\varpi=3\times\tfrac{2}{3}-2\times\tfrac{5}{8}=\tfrac{3}{4},$$

$$\beta_0=\tfrac{2}{5}\times\tfrac{11}{3}\times\tfrac{1}{24}\times3\times2=\tfrac{11}{30},$$

$$\gamma_0=243-\tfrac{512}{9}-44-\tfrac{847}{5}=-27\tfrac{13}{45},$$

illustrating the small value of β_0 and the tendency to negative values of γ_0. In the standard form the curvatures of the end surfaces are

$$\pm\tfrac{1}{2}(j-\kappa)-\frac{1}{2k}\Sigma Q_m k_m,$$

or $$\pm\tfrac{4}{3}-\tfrac{19}{30},$$

or $\frac{7}{10}$ and $-\frac{59}{30}$ respectively. The equation for S, the curvature addition required to give freedom from spherical aberration, is

$$-27\tfrac{13}{45}+10S^2-\tfrac{41}{30}-14S+3+\tfrac{3}{2}=0,$$

which has the two solutions

$$S=\frac{7}{10}\pm\frac{8\sqrt{41}}{30}.$$

The first solution corresponds to the well-known form in which the curvature of the last surface is small. When the lens is reversed so that the parallel beam of light falls on the second component, the signs of the third and fourth terms must be changed. The solutions for this case are therefore

$$S=-\frac{7}{10}\pm\frac{2\sqrt{590}}{30}.$$

The extent to which the sine condition is not satisfied is measured by the value of the left side of (51). In the four cases just considered these values are, with lens 1 leading,

$$\frac{1\pm14\sqrt{41}}{15},$$

and with lens 2 leading,

$$\frac{20\pm7\sqrt{590}}{30},$$

all large errors with the signs differing in the two members of each pair.

Consider next the removal of this aberration by the introduction of an air-gap between the two components. The conditions to be satisfied are

$$\gamma=5+2\varpi$$

and $$\beta=\pm(2+\varpi),$$

the sign depending upon whether the first or the second component is met by the parallel

[1] See *Proc. Phys. Soc. Lond.* xxx. 31.

beam of light. Instead of making use of special formulae for this case it is simpler to solve from (43) and (49) for $S_1\kappa_1$ and $S_2\kappa_2$. The first condition gives

$$28(S_1\kappa_1 - \tfrac{10}{7})^2 - 18(S_2\kappa_2 - \tfrac{13}{6})^2 + 182\tfrac{59}{126} = 0,$$

and the second

$$40S_1\kappa_1 - 26S_2\kappa_2 - 1 = \pm 11,$$

and the solutions are, with lens 1 leading,

$$\left.\begin{matrix} S_1\kappa_1 = 5{\cdot}841 \\ S_2\kappa_2 = 8{\cdot}524 \end{matrix}\right\} \text{ and } \left.\begin{matrix} S_1\kappa_1 = -1{\cdot}351 \\ S_2\kappa_2 = -2{\cdot}541 \end{matrix}\right\},$$

and with lens 2 leading,

$$\left.\begin{matrix} S_1\kappa_1 = 4{\cdot}232 \\ S_2\kappa_2 = 6{\cdot}896 \end{matrix}\right\} \text{ and } \left.\begin{matrix} S_1\kappa_1 = -2{\cdot}952 \\ S_2\kappa_2 = -4{\cdot}157 \end{matrix}\right\}.$$

Finally, cemented objectives built up of three lenses may be considered. From either of the formulae given the values

$$A = -\frac{43}{18}, \quad A' = \frac{101}{9}$$

may be found. The equation from which S has been eliminated may be used to find $1 - 2\theta$ directly. The general equation with the first lens divided is

$$\{(\gamma_0 - A')(1+\varpi)^2 + \beta_0^2(1+2\varpi)\}(1-2\theta)^2$$
$$-2(1+2\varpi)(2+\varpi)\frac{1+G}{1-G}\beta_0(1-2\theta)$$
$$+A'(1+\varpi)^2 - \left(\frac{1+G}{1-G}\right)^2 = 0,$$

and the curvature to be added to the standard form is

$$\frac{1}{2(1+\varpi)}\left\{(2+\varpi)\frac{1+G}{1-G} - \beta_0(1-2\theta)\right\},$$

the curvature of the end face of the standard form being of course

$$\pm\tfrac{1}{2}(j-\kappa) + (1-2\theta)(j_1k_2 - j_2k_1)/2k.$$

Applying these expressions to the solution of the present case, the values found are, with the first lens divided,

$$1 - 2\theta = {\cdot}5117,$$

with extreme curvatures,

$$1{\cdot}741 \text{ and } -0{\cdot}925,$$

and $$1 - 2\theta = -{\cdot}5545,$$

with extreme curvatures,

$$2{\cdot}528 \text{ and } -0{\cdot}138.$$

With the second lens divided the solutions are imaginary.

Any other problem in the calculation of thin objectives is solved by a method exactly comparable to that illustrated. For the discussion of the equations to be used in other problems reference must be made to other sources.

The problem undertaken is not theoretically complete. Any actual lens will not be "thin," but what may be termed a "close" lens, the separations between the surfaces being such as to give the required aperture with a suitable additional substance to enable the glass worker to secure surfaces of good figure. The curvatures obtained from the solution in which thicknesses have been neglected require therefore to be modified in consequence of the departure from correction due to the introduction of thicknesses. Consideration has also to be given to the character of the correction which it is possible to achieve, on which the mathematical conditions that should be obeyed must depend. In general it is found that the first order aberration must not be entirely removed, as the errors of higher orders tend to produce over-correction. The type of correction exhibited by a well-designed objective thus shows small positive aberration, which rises to a maximum and falls to zero for rays which enter near the margin of the aperture, and the procedure adopted must secure the satisfaction of this condition. It is found on investigation that for relative apertures such as are normally employed this condition is satisfied by making no alteration in the curvatures of the surfaces, with the exception of the small change in a shallow surface necessary for the attainment of good colour correction. Owing to this fortunate coincidence the calculation of telescope objectives by the algebraic method described is much to be preferred to the more laborious trigonometrical methods.

§ (12) SYSTEMS WITH MORE THAN TWO GLASSES.—A problem which arises much less frequently than any of those already discussed relates to the calculation of systems in which more than two different glasses are used. When there is one lens of each of n glasses they can be arranged in $n!$ orders affording $\tfrac{1}{2}n!$ cases to be investigated if a complete examination is to be made into the properties of the various systems. In accordance with what has been said of thin cemented systems in general, the most important problem is the determination of the values that γ_0 may take, and particularly the effect on this quantity of variations in the order. It is not difficult to show that when there are three separate glasses the extreme values of γ_0 are obtained by putting a lens of extreme index in the middle, whether the value is the highest or the lowest depending upon whether an odd or an even number of components differ in sign from the complete lens. When there are four components extreme values usually result by placing all the positive lenses together in the order of their indices, and all the negative lenses together also in order. Such complex systems usually arise from

attempts to reduce second order chromatic aberrations.[1]

§ (13) SYSTEMS IN WHICH CHROMATIC ABERRATION IS NOT IMPORTANT.—Systems are occasionally required for use in circumstances where chromatic differences may be neglected, and it is thus possible to construct them using only one kind of glass. Lenses so constructed afford good examples of one or two principles of importance, and they are of practical interest since they may be made to have very small outstanding aberrations for even exceptionally large relative apertures.

When a single lens is employed to form an image of an object placed so that the paraxial magnification is G, the spherical aberration vanishes when

$$\frac{1}{(1-\varpi)^2}+4(1+2\varpi)S^2-8(1+\varpi)S\frac{1+G}{1-G}+(3+2\varpi)\left(\frac{1+G}{1-G}\right)^2=0,$$

and when this cannot be made to vanish the left side may be put in the form

$$\frac{1}{(1-\varpi)^2}-\frac{1}{1+2\varpi}\left(\frac{1+G}{1-G}\right)^2+\frac{4}{1+2\varpi}\left\{(1+2\varpi)S-(1+\varpi)\frac{1+G}{1-G}\right\}^2,$$

which attains its minimum value when

$$S=\frac{1+\varpi}{1+2\varpi}\cdot\frac{1+G}{1-G}.$$

It is usually said that at this position the lens acts like a prism at minimum deviation, the air or glass angles between the ray and the two normals being equal. This is incorrect; for to the order of accuracy here required the ratio of the sine of the angle at the first surface to that at the second is

$$-\frac{2S-(1+G)/(1-G)+1/(1-\varpi)}{2S-(1+G)/(1-G)-1/(1-\varpi)},$$

and when S has the value for which the spherical aberration is a minimum this ratio becomes

$$\frac{2\mu+1-3G}{3-G(2\mu+1)},$$

which shows that the supposed law is only correct in the special case $G=-1$, for which the system is symmetrical. Since $(1+2\varpi)/(1-\varpi)^2$ exceeds unity for all positive values of ϖ less than 4, there is always positive spherical aberration for all values of $(1+G)/(1-G)$ between the limits ± 1, that is, for all real images of real objects. It is possible, however, to eliminate this aberration by substituting for the single lens a number of thin lenses of the same glass in contact. For such a combination (43) gives

$$\beta\kappa^2=2(1+\varpi)\Sigma S_m\kappa_m^2,$$

and (49)

$$\begin{aligned}\gamma\kappa^3=\frac{1}{(1-\varpi)^2}\Sigma\kappa_m^3&+4(1+2\varpi)\Sigma S_m^2\kappa_m^3\\&-8(1+\varpi)\Sigma S_m\kappa_m^2(\kappa_{1,m}-\kappa_{m,n})\\&+(3+2\varpi)\Sigma\kappa_m(\kappa_{1,m}-\kappa_{m,n})^2,\end{aligned}$$

so that the spherical aberration at magnification G is measured by

$$\begin{aligned}\frac{1}{\kappa^3}\Bigg[\frac{1}{(1-\varpi)^2}\Sigma\kappa_m^3&+\frac{4}{1+2\varpi}\Sigma\Bigg\{(1+2\varpi)S_m\kappa_m\\&-(1+\varpi)\left(\kappa_{1,m}-\kappa_{m,n}+\kappa\frac{1+G}{1-G}\right)\Bigg\}^2\kappa_m\\&-\frac{\kappa^3-\Sigma\kappa_m^3}{3(1+2\varpi)}-\frac{\kappa^3}{(1+2\varpi)}\left(\frac{1+G}{1-G}\right)^2\Bigg].\end{aligned}$$

Assuming that the powers of all the component lenses have the same sign as κ, the minimum value of this expression is

$$\frac{1}{3(1+2\varpi)}\left[\left(\frac{2+\varpi}{1-\varpi}\right)^2\frac{\Sigma\kappa_m^3}{\kappa^3}-1-3\left(\frac{1+G}{1-G}\right)^2\right],$$

and the first coefficient being large compared with the other terms, removal of the aberration will be best effected by making $\Sigma\kappa_m^3$ as small as possible compared with κ, that is, by making the power of every component equal. If the number of elements is given, say n, the aberration is removable for all values of ϖ below that which satisfies

$$\frac{1}{n^2}\left(\frac{2+\varpi}{1-\varpi}\right)^2-1-3\left(\frac{1+G}{1-G}\right)^2=0.$$

The smallest number of components for correction is evidently obtained for distant objects, and the greatest number is necessary when the object and image are equal in size. If $G=0$ is taken as a convenient reference case, twice as many components will be required for the same value of ϖ when $G=-1$ and approximately 1·4 times the number when it is about $-\frac{1}{11}$. For $G=0$ the lowest refractive index that may be employed is evidently

$$\mu=\frac{2n+1}{2(n-1)},$$

special values being $\mu>\frac{5}{2}$ for $n=2$, $\mu>\frac{7}{4}$ for $n=3$, $\mu>\frac{3}{2}$ for $n=4$, and $\mu>\frac{11}{8}$ for $n=5$. For the limiting index all the S's are positive and increasing from the first member of the series to the last, the system assuming the form of a series of meniscus lenses in contact on the axis with air gaps between them towards their edges, the curvature differences of the air gaps being equal.

It will usually be necessary for a lens of this kind to satisfy the sine condition in addition to being free from spherical aberration.

[1] *Trans. Opt. Soc.* xxii. 111.

Substituting in the expression for β the value of $S_m\kappa_m$ which results in minimum spherical aberration, it is found that

$$\beta=2\frac{(1+\varpi)^2}{1+2\varpi}\cdot\frac{1+G}{1-G}$$

or $$\beta-(2+\varpi)\frac{1+G}{1-G}+\frac{\varpi}{1+2\varpi}\,\frac{1+G}{1-G}=0,$$

and the sine condition is only fulfilled when the object and image are equal. Evidently this must be met by an increase in the minimum refractive index for a given number of components. Let $S_m\kappa_m$ be given the value

$$v_m+\frac{1+\varpi}{1+2\varpi}(\kappa_{1,m}-\kappa_{m,n})+\frac{2+\varpi}{2(1+\varpi)}\kappa\frac{1+G}{1-G},$$

where v_m is a variable whose value is at present undetermined. The sine condition is satisfied if

$$\Sigma v_m\kappa_m=0,$$

and the spherical aberration coefficient becomes

$$\frac{1}{3(1+2\varpi)}\left\{\left(\frac{2+\varpi}{1-\varpi}\right)^2\frac{\Sigma\kappa_m^3}{\kappa^3}-1\right\}-\frac{1}{(1+\varpi)^2}\left(\frac{1+G}{1-G}\right)^2$$
$$+4(1+2\varpi)\Sigma\frac{v_m^2\kappa_m}{\kappa^3}+\frac{4\varpi}{1+\varpi}\cdot\frac{1+G}{1-G}\Sigma\frac{v_m\kappa_m}{\kappa^2},$$

and the lowest value of the index will be obtained by making all the v's zero and all the powers equal. The equation from which the limit is to be found is

$$\frac{1}{n^2}\left(\frac{2+\varpi}{1-\varpi}\right)^2-1-\frac{3(1+2\varpi)}{(1+\varpi)^2}\left(\frac{1+G}{1-G}\right)^2=0,$$

which gives when $G=0$ and $n=4$ the approximate solution $\varpi=\cdot648$ or $\mu=1\cdot543$.

If the lens determined by this approximate solution has thicknesses inserted without change in the curvatures, the spherical aberration will be very slightly over-corrected, and the departure from the sine condition appreciably over-corrected. The latter is remedied by increasing all the curvatures by a common value, making the lenses more decidedly meniscus, but in so doing the spherical aberration will be over-corrected. This is remedied by a further increase in refractive index, solving for the minimum spherical aberration and suitable under-correction of the sine condition. For a relative aperture of about $f/1$ a suitable refractive index is approximately 1·6. The zonal variations in the spherical aberration and in the sine ratio for even so large an aperture as this will be found extremely small.

§ (14) More Complex Systems.—It is not possible in the space available for this article to enter on a discussion of the way in which the principles applied above in considering thin objectives may be extended so as to cover systems of complex construction for large apertures such as microscope objectives, or photographic lenses where large apertures are required to be combined with a large field of view. It must suffice to say that, by the aid of theorems which assume a very neat form, it is possible to determine the conditions which each lens should satisfy on the assumption that it may be regarded as a thin lens finitely separated from its neighbours, and that a definite treatment applied to this simplified lens will yield the correct data for the real lens. The importance of this procedure lies in its automatic evasion of attempts at the theoretically impossible, to which pursuit the designer without some such guide is very apt to devote a great part of his time. Indeed little experience of lens calculations is required to force home the conviction that almost every attractive scheme for the removal of aberrations involves direct conflict with Fermat's principle. Before concluding, it is, however, desirable to mention briefly a few general results which have many applications.

It was shown in equations (15) that the resolved equations governing the passage of light through the system are of the form

$$\mu_m M_m-\mu_{m-1}M_{m-1}+\eta_m K_m=0,$$
$$\eta_{m+1}-\eta_m-\mu_m M_m\tau_m\varpi_m=0.$$

Let another set of such quantities be distinguished by the introduction of accents. Then, whatever meaning these quantities may bear, the value of

$$\Sigma\eta_m(\mu'_m M'_m-\mu'_{m-1}M'_{m-1}+\eta'_m K'_m)$$
$$-\Sigma\mu_m M_m(\eta'_{m+1}-\eta'_m-\mu'_m M'_m\tau'_m\varpi'_m)$$
$$-\Sigma\eta'_m(\mu_m M_m-\mu_{m-1}M_{m-1}+\eta_m K_m)$$
$$+\Sigma\mu'_m M'_m(\eta_{m+1}-\eta_m-\mu_m M_m\tau_m\varpi_m)$$

is identically zero. In other words,

$$\sum_1^n\eta_m\eta'_m(K'_m-K_m)$$
$$+\sum_1^{n-1}\mu_m M_m\mu'_m M'_m(\tau'_m\varpi'_m-\tau_m\varpi_m)$$
$$=\eta'_n\mu_n M_n-\eta_n\mu'_n M'_n-\eta'_1\mu_0 M_0+\eta_1\mu'_0 M'_0,$$

or

$$\sum_1^n\eta_m\eta'_m(K'_m-K_m)$$
$$+\sum_0^n\mu_m M_m\mu'_m M'_m(\tau'_m\varpi'_m-\tau_m\varpi_m)$$
$$=y'_n\mu_n M_n-y_n\mu'_n M'_n-y'_0\mu_0 M_0+y_0\mu'_0 M'_0, \quad (61)$$

where y_0, y_n, y'_0, y'_n are arbitrary points in the external media on the two rays, distant τ_0, τ_n, τ'_0, τ'_n from the points in which the first and last surfaces are encountered.

These expressions give at once the conditions that paraxial rays relating to light of different colours should form images free from colour. Special results of importance are obtained by considering the case in which both rays are initially parallel to the axis, so that

$M_0 = M'_0 = 0$ and y_n and y'_n are made to vanish by selecting as the final end points the intersections of the ray with the axis. As the rays are paraxial $\tau_m = \tau'_m = t_m$, and since it may be assumed that the difference of index is small enough for the rays never to become appreciably separated if they are initially coincident, the second expression gives

$$\sum_1^n \eta_m{}^2 \delta\kappa_m + \sum_1^{n-1} \mu_m{}^2 M_m{}^2 t_m \delta\varpi_m + M_n{}^2 \delta v_n = 0,$$

where the last medium is non-dispersive and of unit index. It is permissible to divide throughout by $\eta_1{}^2$ and write for the η's and M's their values in terms of the direct paraxial ray, with the result that

$$\kappa^2 \delta v_n + \sum_1^n \left(\frac{\partial\kappa_{1,m}}{\partial\kappa_m}\right)^2 \delta\kappa_m + \sum_1^{n-1} t_m \kappa_{1,m}{}^2 \delta\varpi_m = 0$$

gives the change δv_n in the position of the second principal focus as the colour changes. The corresponding expression for the first principal focus may be written down by analogy. To find the variation of the power, assume $M_0 = M'_n = 0$. The right side of (61) is then $y_0 y'_n(K' - K)$. Dividing throughout by $y_0 y'_n$ and neglecting differences between the colours in the coefficients leads to

$$\delta\kappa = \sum_1^n \frac{\partial\kappa_{1,m}}{\partial\kappa_m} \frac{\partial\kappa_{m,n}}{\partial\kappa_m} \delta\kappa_m - \sum_1^{n-1} t_m \kappa_{1,m} \kappa_{m+1,n} \delta\varpi_m.$$

§ (15) Chromatic Aberration.—When the system is a thin lens the equations reduce to the single condition

$$\sum_1^n \delta\kappa_m = 0$$

for freedom from all chromatic aberrations for paraxial rays, as would be inferred from the fact that when κ is known the paraxial rays are all determined. For a discussion of the properties of glass as regards variation of index for colour, reference should be made to the article on "Optical Glass." For present purposes it will suffice to mention that the principal characteristic of any glass is given by a coefficient ν which gives the difference of power for the C and F lines of hydrogen in terms of that for the D sodium lines in the form

$$\kappa_F - \kappa_C = \frac{\kappa_D}{\nu},$$

and the rays for the C and F lines will thus be refracted alike if

$$\Sigma \frac{\kappa_D}{\nu} = 0.$$

When this condition is satisfied the outstanding chromatic aberrations are those generally considered most satisfactory for images to be examined visually. For photographic use it is usual to secure agreement between the D and G′ lines.

The condition for a thin doublet lens is automatically satisfied by giving κ_1 and κ_2 the values

$$\kappa_1 = \nu_1 \frac{\kappa}{\nu_1 - \nu_2},$$

$$\kappa_2 = -\nu_2 \frac{\kappa}{\nu_1 - \nu_2}.$$

When three glasses are employed it is generally expected to remove the second order differences of colour, and the solution of the problem is easy when once suitable glasses have been chosen. The properties of all glasses hitherto manufactured approximate very closely to a law which results in the lens having zero power when the appropriate conditions are satisfied. It is easy to see that a higher degree of correction is attained if the glasses vary in such a way that when the changes of index between two pairs of reference spectrum lines are chosen as variables, the points representing the glasses when these are plotted in perpendicular directions are collinear. For manufactured glasses, however, these partial ratios are practically linear functions of the ν's with the consequence stated above. A way out of the difficulty is afforded by the use of transparent crystalline materials, of which fluorite is much the most important. This substance exhibits no double refraction, and its representative point on the chart lies almost on the straight line about which normal glasses group themselves, while the linear relation between the dispersions and the ν values does not hold when fluorite is one medium and glasses are selected for the other two. By means of such a chart all suitable combinations of glasses may be obtained in a few minutes, and when the remaining conditions are specified the field of choice can quickly be reduced to one or two alternative combinations.[1]

§ (16) Aberration Formulae.—Equation (61) interpreted in the broadest way forms a sufficient basis for the construction of formulae for aberrations to any required order of accuracy. Such formulae may be arranged on many plans, but, though they have what many workers will regard as peculiar disadvantages, there is much to be said in favour of a development which remains entirely symmetrical in the variables for the initial and final media. An illustration of a symmetrical development will be afforded in the brief consideration given later to the application of Hamilton's method to optical problems. The preceding discussion of the properties of thin lenses also is symmetrical in character, and the three quantities in terms of which the aberrations have been expressed are capable of immediate application to the system used in the reversed

[1] *Trans. Opt. Soc.* xxii. 99.

direction with as much ease as in the original direction. This is a comparatively minor matter when only first order aberrations are under discussion, but it becomes extremely important when higher order effects are under consideration in order to secure the greatest economy of labour.

In a number of problems in which (61) might be employed it is advantageous to return to first principles. As an example, the formula giving the positions of the primary or tangential focal lines may be taken. Suppose a ray is refracted at the origin on a sphere which has its centre of curvature on the x axis, and assume that the ray lies in the plane $z=0$. Consider the length of the path from a point on the surface near the origin to the point $\rho\cos\phi$, $\rho\sin\phi$, 0. The point on the surface is approximately $\frac{1}{2}R(\eta^2+\zeta^2)$, η, ζ and the distance between the two points is

$$\rho-\eta\sin\phi+\frac{\eta^2}{2}\left(\frac{\cos^2\phi}{\rho}-R\cos\phi\right)+\frac{\zeta^2}{2}\left(\frac{1}{\rho}-R\cos\phi\right)+\ldots,$$

and thus the ray through η, ζ from $\rho\cos\phi$, $\rho\sin\phi$, 0 to $\rho'\cos\phi'$, $\rho'\sin\phi'$, 0 exceeds in optical path the ray through the origin by

$$-\eta(\mu'\sin\phi'-\mu\sin\phi)+\frac{\eta^2}{2}\left(\frac{\mu'\cos^2\phi'}{\rho'}-\frac{\mu\cos^2\phi}{\rho}-K\right)+\frac{\zeta^2}{2}\left(\frac{\mu'}{\rho'}-\frac{\mu}{\rho}-K\right)+\ldots$$

In the ordinary instrument there is no uniform change of path from point to point of the surface, so $\mu'\sin\phi'=\mu\sin\phi$, the ordinary law of refraction, is satisfied, which gives merely the direction in which the wavelets will reinforce one another. The following terms indicate the distances in these directions in which special concentrations will be found corresponding to the foci of geometrical optics. The meaning is obviously that considerable latitude may be given to ζ without producing any appreciable change of path in the positions given by the law hitherto considered exclusively, and similar latitude to η at positions given by the new law

$$\frac{\mu'\cos^2\phi'}{\rho'}-\frac{\mu\cos^2\phi}{\rho}=K. \quad . \quad (62)$$

Since the former law may be derived by considering a single ray alone without reference to neighbouring rays, it is evident that (62) represents the differential of that law, and it will be expected that relations derived from (62) will exhibit this form when compared with those obtained from the earlier law.

A simple application of this law of importance relates to rays passing near the axial point of a thin system of lenses. There are no changes in the angles of incidence between the two surfaces bounding any medium, since all are met where they are normal to the axis, and the equations for the individual surfaces are thus directly additive, giving, since

$$M_0=M_n, \quad N_0=N_n,$$

$$\frac{1}{\rho_n}-\frac{1}{\rho_0}=\frac{K}{1-\frac{1}{4}\{(M_0+M_n)^2+(N_0+N_n)^2\}},$$

so that the change is equivalent to replacing the ϖ in the expression for K, when $\eta=\zeta=0$ by $2+\varpi$. It at once follows that the relation between the curvatures of primary conjugate surfaces corresponding to (52) for the secondary surfaces is

$$\bar{s}-\bar{s}'-(3+\varpi)\kappa=0. \quad . \quad . \quad (63)$$

The interpretation of the two results may be expressed in a variety of ways; one of the simplest, which holds for systems in general and not solely for thin systems, is that the paraxial region of any image surface is most simply referred to a surface whose curvature is less than that of the object surface by $\varpi\kappa$, and that the relative distances of primary and secondary foci from this surface are in the ratio 3 to 1.

It may readily be seen that the defect known as coma, which is a manifestation of failure to satisfy the sine condition, shows a similar ratio of 3 to 1 for errors in and normal to the plane containing the principal ray. Careful consideration will show that simple laws which may at first appear to relate to the larger errors in fact must be interpreted in relation to the smaller, and the corresponding laws for the former will assume differential forms. The results already quoted will become evident in the discussion of Hamilton's method.

§ (17) A System of Thin Lenses.—It has been assumed in discussing the properties of thin lenses that reference to the unit surfaces provides a convenient means of distinguishing one ray from another. When a succession of separated lenses have to be considered, or a stop at a distance from the lens limits the aperture so that rays from different parts of the object meet the lens at different parts of its aperture, it is necessary to adopt a different reference system. The results already established on change of variables give aberration coefficients $\frac{1}{2}\delta_1$, $\frac{1}{2}\delta_2$, $\frac{1}{2}\{3\delta_3+\varpi(S-G)^2\}$, $\frac{1}{2}\{\delta_3+\varpi(S-G)^2\}$, $\frac{1}{2}\delta_4$, $\frac{1}{2}\delta_5$ for spherical aberration, coma, primary curvature, secondary curvature, distortion, and stop aberration respectively, where S is the magnification for the stop in the entire system, G the magnification for the image, and the

quantities $\delta_1, \delta_2 \ldots \delta_5$ are defined by the equations

$$4\delta_1(1-G)^{-4} = \gamma - 4\beta\mathfrak{G} + (3+2\varpi)\mathfrak{G}^2,$$

$$4\delta_2(1-G)^{-3}(1-S)^{-1} = \gamma - \beta(3\mathfrak{G}+\mathfrak{S}) + (1+\varpi)\mathfrak{G}^2 + (2+\varpi)\mathfrak{G}\mathfrak{S},$$

$$4\delta_3(1-G)^{-2}(1-S)^{-2} = \gamma - 2\beta(\mathfrak{G}+\mathfrak{S}) + 2(1+\varpi)\mathfrak{G}\mathfrak{S} + \mathfrak{S}^2,$$

$$4\delta_4(1-G)^{-1}(1-S)^{-3} = \gamma - \beta(\mathfrak{G}+3\mathfrak{S}) + \varpi\mathfrak{G}\mathfrak{S} + (3+\varpi)\mathfrak{S}^2,$$

$$4\delta_5(1-S)^{-4} = \gamma - 4\beta\mathfrak{S} + (3+2\varpi)\mathfrak{S}^2,$$

where $$\mathfrak{G} = \frac{1+G}{1-G}, \quad \mathfrak{S} = \frac{1+S}{1-S}.$$

The fact that the three middle expressions do not transform into one another by the interchange of G and S is of great importance, since it involves very severe restrictions on what may be achieved by systems built up from thin lenses. This property may be regarded as a direct consequence of Fermat's principle.

The expressions that have been obtained for β and γ show that the δ's for a complex system are consistent with the additive laws

$$(f\delta_r)_{1,n} = \Sigma(f\delta_r)_m G_{m+1,n}^{5-r} S_{m+1,n}^{r-1},$$

where f is the focal length of the part determined by the suffix outside the bracket, and $G_{m+1,n}$, $S_{m+1,n}$ denote the magnifications of the image and stop in the portions of the system following lens m; that is to say, they are the continued products of the subsequent magnifications at individual lens or surfaces. These additive laws hold generally, and may be established from first principles without a knowledge of the expressions from which the aberrations are to be numerically computed. As an example, spherical aberration may be taken. The section of the caustic resulting from the presence of this aberration must for the lowest order be of the form of a semi-cubical parabola

$$x^3 = y^2 f\delta,$$

where δ is a numerical coefficient measuring the amount of the aberration, and f is the focal length introduced in order that the coefficient δ may be of zero dimensions. If now a system is divided into two parts 1 and 2, in each of which aberration of this type is present, the coefficient for the complete lens will depend upon the coefficients for the two separate portions, since it must vanish if both of these are zero. If the first part is without aberration the linear coefficient $f\delta$ for the whole will obviously be identical with that for the second portion alone. If, on the other hand, the second part is free from aberration the final image will be the image of the caustic formed after the light has traversed the first part, and this image will be formed according to paraxial laws. Thus if G_2 is the transverse linear magnification for the second part of the system $y_2 = G_2 y_1$, $x_2 = G_2^2 x_1$, and

$$(f\delta_1)_{1,2} = \frac{x_2^3}{y_2^2} = \frac{(G_2^2 x_1)^3}{(G_2 y_1)^2} = G_2^4\frac{x_1^3}{y_1^2} = G_2^4(f\delta_1)_1.$$

Since the errors considered are first order effects, the law of addition is linear, and by combining the two cases the general law must be

$$(f\delta_1)_{1,2} = (f\delta_1)_1 G_2^4 + (f\delta_1)_2,$$

and by repeated application the result already given is established for $r=1$. It will be noted that when the system is divided by surfaces instead of by lenses, so that the refractive index is not necessarily unity at the points of division, the factors require modification.

§ (18) HAMILTON'S METHOD.—Brief consideration must now be given to the very elegant method of investigating problems in reflection and refraction described by Sir W. R. Hamilton before the Royal Irish Academy in a series of papers covering the years 1824 to 1832. Essentially the principle is the use of a potential function of certain variables the partial derivatives of which determine all remaining unknowns in terms of these variables. Theoretically the method is extremely powerful and its application in other fields has proved of extraordinary importance. It is thus particularly noteworthy that in its application in the field for which it was primarily designed it should be generally considered to have failed to justify expectations. One reason doubtless lies in the nature of the subject itself, which requires all problems to be investigated with a minuteness not called for or appreciable in other fields. Another contributory cause is perhaps that Hamilton himself, like most mathematicians, was more interested in the broad problems to which his discovery was applicable, and in deriving general results to which it led with great ease, than in the tedious task of constructing expressions which gave a numerical measure of the effects so readily investigated qualitatively, and that this task has proved a formidable one in less able hands. Whatever the true cause may be, the result remains that no notable use was made of the method in the first fifty years after its publication. Perhaps the best testimony to this failure is that it was largely forgotten that it was to these optical problems that potential methods were first applied, with the result that Bruns in Germany believed that he was breaking entirely new ground in so applying Hamiltonian tools, and published a very extensive and heavy investigation which, for the most part, constituted a duplication of Hamilton's original

work. To the potential function which he used for the greater part of his investigations Hamilton applied the name "Characteristic Function"; but use was also made of another function, with different variables and allied properties, which was designated the "Allied Characteristic Function." Bruns, with characteristic thoroughness, investigated all the functions which could be utilised for the purpose, and proposed for them the generic title of "Eikonal." It is convenient to retain Hamilton's name in the sense in which he used it, and to restrict "Eikonal" to a particular function which, like the characteristic function, is symmetrical in its variables, and takes for these variables the direction cosines of a ray instead of the co-ordinates of a point. In both cases the function has the dimensions of a length, and in the case of the characteristic function represents the length of the optical path between the points whose co-ordinates are taken as variables, and in the eikonal the length of the optical path between the feet of perpendiculars to the ray from two selected points in the object and image spaces respectively. Instead of introducing the full number of variables, since all that is required is the identification of the initial and final portions of a single ray, one of them in each space may be absent, and intersections with fixed planes will be derived.

Assuming that the system is symmetrical about an axis, it is convenient in the case of the characteristic function to adopt y, z and y', z' as variables, the axis of symmetry coinciding with the axes of x and x'. If then V is the optical length of the path between fixed planes expressed as a function of the points in which the ray intersects these planes, the direction cosines of the ray before and after traversing the system are given by

$$\mu M = -\frac{\partial V}{\partial y}, \quad \mu N = -\frac{\partial V}{\partial z},$$

$$\mu' M' = \frac{\partial V}{\partial y'}, \quad \mu' N' = \frac{\partial V}{\partial z'}.$$

Similarly with the eikonal E, if the optical path is expressed as a function of M, N, M′, N′, the reference points being on the axis, the co-ordinates of the points in which the ray meets the planes normal to the axis through these reference points are determined from

$$y = \frac{\partial E}{\partial(\mu M)}, \quad z = \frac{\partial E}{\partial(\mu N)},$$

$$y' = -\frac{\partial E}{\partial(\mu' M')}, \quad z' = -\frac{\partial E}{\partial(\mu' N')}.$$

Since the system is symmetrical about an axis, V will be a function of y^2+z^2, $yy'+zz'$, $y'^2+z'^2$ only, and E a function of M^2+N^2, $MM'+NN'$, $M'^2+N'^2$ only. Put $(1) = \mu^2(M^2+N^2)$, $(2) = \mu\mu'(MM'+NN')$, $(3) = \mu'^2(M'^2+N'^2)$. Then E is a function of (1), (2), (3), and these may be transformed into other variables as found convenient for discussing particular problems. Suppose that it is desired to consider the conditions for freedom from aberration for magnification G, the reference points being the axial points of the object and image planes. Let there be a stop for which the magnification is S. Introduce the variables I, II, III defined by

$$\left.\begin{aligned} &\mathrm{I} \quad (S-G)^2 = (1) - 2G(2) + G^2(3) \\ &\mathrm{II} \quad (S-G)^2 = (1) - (G+S)(2) + GS(3) \\ &\mathrm{III} \quad (S-G)^2 = (1) - 2S(2) + S^2(3) \end{aligned}\right\}, \qquad (64)$$

so that

$$\left.\begin{aligned} (1) &= S^2\mathrm{I} - 2GS\mathrm{II} + G^2\mathrm{III} \\ (2) &= S\mathrm{I} - (G+S)\mathrm{II} + G\mathrm{III} \\ (3) &= \mathrm{I} - 2\mathrm{II} + \mathrm{III} \end{aligned}\right\}; \qquad (65)$$

then, denoting differentiation by a suffix,

$$y(S-G)^2 = (\mu M - G\mu' M')(2E_{\mathrm{I}} + E_{\mathrm{II}}) + (\mu M - S\mu' M')(E_{\mathrm{II}} + 2E_{\mathrm{III}}),$$

$$y'(S-G)^2 = (\mu M - G\mu' M')(2GE_{\mathrm{I}} + SE_{\mathrm{II}}) + (\mu M - S\mu' M')(GE_{\mathrm{II}} + 2SE_{\mathrm{III}}), \qquad (66)$$

so that

$$(y' - Gy)(S-G) = (\mu M - G\mu' M')E_{\mathrm{II}} + 2(\mu M - S\mu' M')E_{\mathrm{III}},$$

and similarly

$$(z' - Gz)(S-G) = (\mu N - G\mu' N')E_{\mathrm{II}} + 2(\mu N - S\mu' N')E_{\mathrm{III}}.$$

Now if the image of the plane object falls on the plane and is free from aberration, $y' - Gy = z' - Gz = 0$ for all rays. Moreover, in general $M/N \neq M'/N'$, since this implies that the rays are restricted to the axial plane containing the object and image points. Consequently E_{II} and E_{III} must vanish identically; that is, E must be a function of I only, and all rays arising from the same object point must have identical values for $\mu M - G\mu' M'$ and $\mu N - G\mu' N'$. This is a generalised form of the well-known sine condition, and reference to (33) shows that it involves the result stated earlier, that K must have the same value for all rays arising from the same point of this object plane.

The sine condition as usually given relates only to rays arising from a point on the axis. Equation (66) shows that in this case $\mu M - G\mu' M'$ must be zero—E_{I} cannot vanish, for its constant term is the focal length of the system. The result may, therefore, be stated in the form that the ratio $\mu M/\mu' M'$ must be constant for all rays through the axial point and equal to the linear magnification for paraxial rays.

The connection between K and E is readily found from (32) and (66),

$$\frac{1}{K}=\frac{2G}{(S-G)^2}E_1=-E_{(2)} \text{ by (64)},$$

and it is easy to verify that this relation is always true. Thus $-E_2$ is the focal length for the ray determined by the values given to the variables. E_1 and E_3 serve to fix the positions of the principal foci on this ray. The relations between the two systems enable the results of calculation made by one of them to be readily stated in terms of the other, and in particular they show that phase relationship at and near the foci may be directly obtained from ray tracing by simple (usually graphical) integration. Notwithstanding the simplicity of direct calculations of phase differences, it will be found that the advantage lies with this graphical method. Incidentally it affords a convincing demonstration of the very small differences in the conclusions to which different theories lead for the location of the best focus in the presence of aberration, and thus justifies the recognised practice of usually disregarding the wave theory in calculating the details of construction of an optical system.

§ (19) The Cosine Condition.—Investigations have been made of the meaning to be attached to definite departures from the sine condition, and the problem is of great importance in economising in the volume of calculations. Some of the conclusions derived are invalidated by erroneous assumptions made in the course of the work, and it will not be superfluous to inquire into the problem here. It is not necessary to assume that the system is symmetrical in any sense, and any reference planes may be chosen for the object and image spaces.

Consider three neighbouring rays (*a*), (*b*), and (*c*) which have respectively the initial and final direction cosines

(*a*)	L,	M,	N;	$L'+\delta L'$,	$M'+\delta M'$,	$N'+\delta N'$:
(*b*)	L,	M,	N;	L',	M',	N':
(*c*)	$L+\delta L$,	$M+\delta M$,	$N+\delta N$;	L',	M',	N':

and which intersect the reference planes in the points

(*a*)	0,	y,	z;	..	..	..:
(*b*)	0,	$y+\delta y$,	$z+\delta z$;	0,	$y'+\delta y'$,	$z'+\delta z'$:
(*c*)	..	..	.. ;	0,	y'	z'.

Add to E the suffixes *a* and *c* when $M'+\delta M'$, $N'+\delta N'$ are to be substituted for M' and N', and $M+\delta M$, $N+\delta N$ for M and N respectively. Then

$$\mu\delta y=\frac{\partial E}{\partial M}-\frac{\partial E_a}{\partial M},$$

and expanding E_a by Taylor's theorem in terms of E and its derivates

$$\mu\delta y=-\frac{\partial^2 E}{\partial M\partial M'}\delta M'-\frac{\partial^2 E}{\partial M\partial N'}\delta N'$$
$$+ \text{ higher powers of } \delta M' \text{ and } \delta N'.$$

Similarly

$$\mu\delta z=-\frac{\partial^2 E}{\partial N\partial M'}\delta M'-\frac{\partial^2 E}{\partial N\partial N'}\delta N'+\ldots;$$

also

$$\mu'\delta y'=-\frac{\partial E}{\partial M'}+\frac{\partial E_c}{\partial M'}$$
$$=\frac{\partial^2 E}{\partial M\partial M'}\delta M+\frac{\partial^2 E}{\partial N\partial M'}\delta N+\ldots,$$

and

$$\mu'\delta z'=\frac{\partial^2 E}{\partial M\partial N'}\delta M+\frac{\partial^2 E}{\partial N\partial N'}\delta N+\ldots$$

Multiply these four expressions by δM, δN, $\delta M'$, $\delta N'$ and add. Then

$$\mu\delta y\delta M+\mu\delta z\delta N+\mu'\delta y'\delta M'+\mu'\delta z'\delta N'=0 \qquad (67)$$

to at least the third order in the small quantities δM, δN, $\delta M'$, $\delta N'$. This relation contains the general result to be established since the directions of the axes are quite arbitrary, but it may be clearer to express it in a somewhat different form.

If rays (*a*) and (*c*) satisfy a relation of the form

$$\cos\theta=p\cos\theta'+q, \qquad (68)$$

where θ and θ' are angles made with fixed straight lines in the object and image spaces respectively having direction cosines l, m, n and l', m', n' respectively, the expression of these two conditions in terms of the direction cosines gives

$$l\delta L+m\delta M+n\delta N+p(l'\delta L'+m'\delta M'+n'\delta N')=0,$$

Now a general point on ray (*a*) is

$$\rho L,\quad y+\rho M,\quad z+\rho N,$$

and if this ray is displaced without rotation through a distance d in the direction l, m, n a point on the displaced ray is

$$\rho L+dl,\quad y+\rho M+dm,\quad z+\rho N+dn,$$

and a particular point is

$$0,\quad y+\frac{d}{L}(Lm-Ml),\quad z+\frac{d}{L}(Ln-Nl),$$

and if this displaced ray is identical with (*b*)

$$\delta y=\frac{d}{L}(Lm-Ml),\quad \delta z=\frac{d}{L}(Ln-Nl).$$

Similarly if ray (*c*) coincides with (*b*) when displaced through a distance d' without rotation

$$\delta y'=\frac{d'}{L'}(L'm'-M'l'),\quad \delta z'=\frac{d'}{L'}(L'n'-N'l').$$

Substitute these values in (67). Then

$$\mu d(l\delta L+m\delta M+n\delta N)$$
$$+\mu'd'(l'\delta L'+m'\delta M'+n'\delta N')$$
$$=\frac{\mu dl}{L}(L\delta L+M\delta M+N\delta N)$$
$$+\frac{\mu'd'l'}{L'}(L'\delta L'+M'\delta M'+N'\delta N')$$
$$+\text{third order quantities in } \delta M, \delta N, \delta M', \delta N'.$$

It follows that if these quantities are small all rays selected by obedience to the condition (68) will, if displaced in the object space a small distance d parallel to the first direction, be displaced through the distance $p\cdot\mu d/\mu'$ parallel to the second direction in the image space. As the condition will in general

define an object and an image caustic, these caustics may be displaced without rotation or deformation in the given directions, provided the ratio of the displacements has the required value.

The cosine condition (68), which has been established without any assumption relative to the aberrations present, includes as particular cases the various theorems described previously. If the system is symmetrical about an axis, the rays from points at the same distance from the axis in a given object plane will be refracted similarly, and thus if $z=z'=0$, $(\delta z'/\delta z)=(y'/y)$. For such a case equation (33) may be applied, showing that $\mu M/\mu' M'$ is the magnification perpendicular to the plane containing the ray, and $\mu \delta M/\mu' \delta M'$ is by (67)[1] the magnification in that plane.

§ (20) Magnification for Displacement along the Axis.—Another case of importance relates to the magnification for displacement along the axis. If the axis itself is included as one of the rays of the group to be considered, q must have the value $1-p$, and the axial magnification is $\mu p/\mu'$, where $p=(1-\cos\theta)/(1-\cos\theta')=(\sin\frac{1}{2}\theta/\sin\frac{1}{2}\theta')^2$, and for paraxial rays this becomes $(\mu' G/\mu)^2$, where G is the transverse magnification. If, therefore, no change is to be made in the spherical aberration for a small axial displacement the condition to be satisfied is

$$\frac{\mu\sin(\theta/2)}{\mu'\sin(\theta'/2)}=G.$$

It has been found that when there is freedom from aberrations at magnification G, E is an arbitrary function of I. This evidently implies that there is one degree of freedom remaining in the way in which the aberrations of every order are corrected, and this freedom may be interpreted in terms of the spherical aberration in the image of the stop. If these aberrations are assigned in addition to the position of the stop the properties of the entire system are fully determined. The most important case is when $S=1$ and the stop is free from spherical aberration. In this case

$$E\kappa=-\frac{(1-G)^2}{G}\sqrt{1-I};$$

for if the reference planes are moved to the stop planes E becomes E′, where

$$E'\kappa=-\frac{(1-G)^2}{G}\sqrt{1-I}+\frac{1-G}{G}\sqrt{1-(1)}-(1-G)\sqrt{1-(3)},$$

and when $M=M'$, $N=N'$ this gives

$$E'=y=y'=z=z'=0$$

for all rays. It will be observed that there are here three points on the axis for which spherical aberration of all orders is removed—one for magnification G and two coincident points for magnification unity. The equation for freedom from spherical aberration of any order is necessarily of even order in the magnification, so there is at least one other point where the aberration of any assigned order disappears. In general there is no other position for which the aberration vanishes for all orders. The first order aberration, for example, vanishes for the magnification

$$-\frac{1-2G}{2-G},$$

and the corresponding value for $E\kappa$ is

$$-\frac{(1-G)^2}{G}\sqrt{1-I}+\frac{1-G^2}{(1-2G)G}\sqrt{1-(1)}+\frac{1-G^2}{2-G}\sqrt{1-(3)},$$

from which it readily follows that spherical aberration of higher orders is present.

§ (21) Standard Forms for Aberration Expansions.—The fact that entire freedom from aberrations for magnification G requires E to be a function of I only shows that all terms which have $(MN'-M'N)$ as a factor must vanish. This enables a number of conditions to be written down without reference to the magnification, and these conditions will be found to restrict severely the types of system that can satisfy the conditions. The condition for the first order aberrations, for instance, is $\varpi=0$, and this condition has been associated with the name of Petzval. The difficulty of satisfying this and other conditions simultaneously is illustrated by the general experience that in even the most highly corrected systems a compromise is necessary in consequence of which this condition is usually far from satisfied. The example given of a case in which E is expressed in a finite form suggests that in systems which are well corrected there should be no question about the convergency of E or of its derivates when expanded in a series of ascending powers of (1), (2), and (3). It follows that the evaluation of the coefficients of the terms in such an expansion should in such a system give a reliable measure of the outstanding aberrations. The converse, however, may in some instances not hold; it is quite conceivable that a system which has small values for the coefficients of the terms of low orders may not be satisfactory through the corresponding expansion for a part of the system being divergent. It is therefore important that systems of novel construction calculated by means of the expression for these coefficients should be checked in a way which does not assume the validity of such an expansion. Such a check is furnished by tracing selected rays through the system and comparing the results with those found from the earlier work.

[1] It should be particularly noted that for use in the ordinary notation the third and fourth terms of (67) must have their signs changed, because the differentials of the direction cosines belong to one ray for the object space and the other ray for the image space.

It is convenient to have a standard form for such expansions, and it is highly advantageous to keep it symmetrical in form. For this reason, instead of choosing as the reference planes a pair of object and image planes, the planes normal to the axis through the principal foci are selected. Apart from an additive constant, which is of no interest, E then takes the standard expanded form

$$E\kappa = -(2) - \tfrac{1}{8}[A'(1)^2 - 4B'(1)(2) + 4C(2)^2 + 2(C+\varpi)(1)(3) - 4B(2)(3) + A(3)^2] - \text{terms of higher orders.}$$

Evidently if the light is reversed in direction, to secure similar results it is only necessary to interchange A and A′, B and B′. From their definitions C and ϖ will be symmetrical with respect to the two directions. Assuming that the external media are of unit index, the value of E when the reference points are in the planes for magnification G is found by adding $-(1/G)\sqrt{1-(1)} - G\sqrt{1-(3)}$. Expanding and substituting from (65) gives

$$E\kappa = \frac{(S-G)^2}{2G}\mathrm{I} - \tfrac{1}{8}\Big[\delta_1\mathrm{III}^2 - 4\delta_2\mathrm{II\,III} + 4\delta_3\mathrm{II}^2 + 2\{\delta_3 + \varpi(S-G)^2\}\mathrm{I\,III} - 4\delta_4\mathrm{I\,II} + \Big\{\delta_5 - \frac{(S-G)(S^3-G)}{G}\Big\}\mathrm{I}^2\Big] \ldots,$$

where

$$\delta_1 = A - G(4B+1) + 2G^2(3C+\varpi) - G^3(4B'+1) + G^4A',$$

$$\delta_2 = A - G(3B+1) + G^2(3C+\varpi) - G^3B' - S\{B - G(3C+\varpi) + G^2(3B'+1) - G^3A'\},$$

$$\delta_3 = A - G(2B+1) + G^2C - 2S\{B - G(2C+\varpi) + G^2B'\} + S^2\{C - G(2B'+1) + G^2A'\},$$

$$\delta_4 = A - G(B+1) - S\{3B - G(3C+\varpi)\} + S^2\{3C + \varpi - 3GB'\} - S^3\{B'+1-GA'\}$$

$$\delta_5 = A - S(4B+1) + 2S^2(3C+\varpi) - S^3(4B'+1) + S^4A'.$$

Alternatively, substitution from (64) gives

$$A = G - G^3\frac{S^3-G}{(S-G)^3} + 2\varpi\frac{S^2G^2}{(S-G)^3} + \frac{(\delta_1, \delta_2, \delta_3, \delta_4, \delta_5)(S-G)^4}{(S-G)^4},$$

$$B = -G^2\frac{S^3-G}{(S-G)^3} + \varpi\frac{SG(S+G)}{(S-G)^2} + \frac{(\delta_1-\delta_2, \delta_2-\delta_3, \delta_3-\delta_4, \delta_4-\delta_5)(S-G)^3}{(S-G)^4},$$

$$C = -G\frac{S^3-G}{(S-G)^3} + 2\varpi\frac{SG}{(S-G)^2} + \frac{(\delta_1-2\delta_2+\delta_3, \delta_2-2\delta_3+\delta_4, \delta_3-2\delta_4+\delta_5)(S-G)^2}{(S-G)^4},$$

$$B' = -\frac{S^3-G}{(S-G)^3} + \varpi\frac{S+G}{(S-G)^2} + \frac{(\delta_1-3\delta_2+3\delta_3-\delta_4, \delta_2-3\delta_3+3\delta_4-\delta_5)(S-G)}{(S-G)^4},$$

$$A' = \frac{1}{G} - \frac{S^3-G}{(S-G)^3} + \varpi\frac{1}{(S-G)^2} + \frac{\delta_1 - 4\delta_2 + 6\delta_3 - 4\delta_4 + \delta_5}{(S-G)^4},$$

and the same method readily expresses the aberrations when S and G are given in terms of the aberrations for other values S′ and G′.

Comparison of δ_1 and δ_5 shows that the meaning of the latter with relation to the stop corresponds to the meaning of the former for the object. Differentiation shows that

$$M - GM' \doteqdot y'\kappa, \quad N - GN' \doteqdot z'\kappa$$

in accordance with what has been found in (32). The first term gives the well-known paraxial laws, and the remaining terms correspond to aberrations. Writing η' and ζ' for points in the final stop image, which will be given by

$$M - SM' = \eta'\kappa, \quad N - SN' = \zeta'\kappa,$$

the image defects are given approximately by

$$y' - Gy = -\frac{\kappa^2}{2(S-G)^3}[\delta_1\eta'(\eta'^2+\zeta'^2) - \delta_2\{2\eta'(\eta'y'+\zeta'z') + y'(\eta'^2+\zeta'^2)\} + \delta_3\{\eta'(y'^2+z'^2) + 2y'(\eta'y'+\zeta'z')\} + \varpi(S-G)^2\eta'(y'^2+z'^2) - \delta_4 y'(y'^2+z'^2)],$$

with a corresponding expression for $z' - Gz$. Assuming that a point in the object plane $z=0$ is under consideration, z' may be replaced by zero on the right, giving

$$y' - Gy = -\frac{\kappa^2}{2(S-G)^3}[\delta_1\eta'(\eta'^2+\zeta'^2) - \delta_2 y'(3\eta'^2+\zeta'^2) + \{3\delta_3 + \varpi(S-G)^2\}\eta' y'^2 - \delta_4 y'^3],$$

$$z' = -\frac{\kappa^2}{2(S-G)^3}[\delta_1\zeta'(\eta'^2+\zeta'^2) - \delta_2 2y'\eta'\zeta' + \{\delta_3 + \varpi(S-G)^2\}\zeta' y'^2],$$

from which the character of the various aberrations is readily seen. All are familiar with the exception of that involving δ_2, the coma aberration, which is generally regarded as a rather mysterious effect. The displacements are directly proportional to the distance of the image point from the axis and to the square of the distance from the axis at which the aperture is traversed. The introduction of polar co-ordinates at the stop shows the character of the defect more readily, the y and z displacements being then proportional to

$$y'(\eta'^2+\zeta'^2)(2+\cos 2\theta), \text{ and } y'(\eta'^2+\zeta'^2)\sin 2\theta$$

respectively. A detailed discussion of this aberration will be found in the text-books on advanced geometrical optics.

§ (22) Aberration Coefficients found by the Hamiltonian Method.—The comparison of the curvature terms with those found earlier shows that the characteristic relations for thin lenses in air are

$$A - 2B + C = 1, \quad B - 2C + B' = \varpi, \quad C - 2B' + A' = 1,$$

and further comparison with the terms in δ_1 and δ_2 gives for this case

$$\beta = B' - B, \quad \gamma = 4C + 2\varpi + 1.$$

The Hamiltonian methods are sufficient in themselves for the determination of the aberra-

tion coefficients, by building up those for a compound system in terms of the elements. If the system is divided at the foot of the perpendicular to the ray from an intermediate axial point, it follows from the definition of E as a length that

$$\mathrm{E}-\mathrm{E}_1-\mathrm{E}_2=0.$$

Moreover, if the ray meets the plane through this intermediate point in y, z, these are given by the equations

$$y=-\frac{\partial \mathrm{E}_1}{\partial(\mu \mathrm{M})}, \quad z=-\frac{\partial \mathrm{E}_1}{\partial(\mu \mathrm{N})},$$

as well as

$$y=\frac{\partial \mathrm{E}_2}{\partial(\mu \mathrm{M})}, \quad z=\frac{\partial \mathrm{E}_2}{\partial(\mu \mathrm{N})},$$

and therefore, since E is a function of other variables than M and N, which concern the intermediate medium alone,

$$\frac{\partial}{\partial(\mu \mathrm{M})}(\mathrm{E}-\mathrm{E}_1-\mathrm{E}_2)=\frac{\partial}{\partial(\mu \mathrm{N})}(\mathrm{E}-\mathrm{E}_1-\mathrm{E}_2)=0,$$

and these give two equations for μM and μN in terms of the external variables. A similar process may, if desired, be applied to the external media. The substitution of the values for μM and μN given by these equations in the equation for E solves the problem. The process is tedious rather than difficult, and one reason why other methods have been more successful in practice at once becomes apparent—the equations present themselves in a form requiring reduction, whereas other methods give them directly in their lowest terms.

The same process may be applied to the characteristic function, and indeed the forms of the two expressions are very similar, and when one is known the other can be written down at once. They each have special advantages as well as special disadvantages. The eikonal is particularly convenient in dealing with problems concerning thin lenses, but fails when the system becomes telescopic. If the eikonal is regarded as built upon the power κ, a symmetrical construction which remains finite for a telescope may be founded upon $\partial^2\kappa/\partial\kappa_1\partial\kappa_n$, but this breaks down in the case of a thin lens. The advantage rests with the system that has been followed here, as it is never essential to deal with a telescope as a unit, but a suitable division of the system may always be made in which the two parts are of finite focal length and the aberrations of which require to be suitably related.

A convenient system of equations for the calculation of the coefficients of the eikonal is

$$\mathrm{A}=\Sigma g^2 h, \qquad \mathrm{A}'=\Sigma g'^2 h',$$

$$\mathrm{B}=\Sigma g g' h, \qquad \mathrm{B}'=\Sigma g g' h',$$

$$\mathrm{C}=\Sigma g'^2 h=\Sigma g^2 h', \qquad \varpi=\frac{1}{\kappa}\Sigma\frac{\kappa_m}{\mu_{m-1}\mu_m},$$

where the summations extend over all surfaces, and

$$g_m=\mu_{m-1}\mathrm{R}_m\frac{\partial\kappa_{1,m}}{\partial\kappa_m}-\kappa_{1,m-1}=\mu_m\mathrm{R}_m\frac{\partial\kappa_{1,m}}{\partial\kappa_m}-\kappa_{1,m},$$

$$g'_m=\mu_{m-1}\mathrm{R}_m\frac{\partial\kappa_{m,n}}{\partial\kappa_m}+\kappa_{m,n}=\mu_m\mathrm{R}_m\frac{\partial\kappa_{m,n}}{\partial\kappa_m}+\kappa_{m+1,n},$$

$$h_m=\frac{\partial\kappa_{1,m}}{\partial\kappa_m}\left(\frac{\kappa_{1,m}}{\mu_m^2}-\frac{\kappa_{1,m-1}}{\mu_{m-1}^2}\right),$$

$$h'_m=\frac{\partial\kappa_{m,n}}{\partial\kappa_m}\left(\frac{\kappa_{m,n}}{\mu_{m-1}^2}-\frac{\kappa_{m+1,n}}{\mu_m^2}\right).$$

§ (23) Method of checking the Results.—In checking the reliability of the conclusions reached by the approximate methods that have been described, it is convenient to have the terms representing aberrational effects separated from those which are in accordance with paraxial laws. As a rule it is only necessary to consider rays which lie in a plane containing the axis, and these rays may be rapidly traced by means of the formulae

$$\sin\phi=\sin\psi+\mathfrak{h}\mathrm{R},$$

$$\mu'\sin\phi'=\mu\sin\phi,$$

$$\mathfrak{h}'-\mathfrak{h}=\frac{4\mathfrak{h}(\sin\phi-\sin\phi')(\sin\phi'+\sin\psi)}{\sin^2\psi_0'+(\cos\psi+\cos\phi+\cos\phi')^2} \quad 1,$$

$$\sin\psi'=\sin\psi_0'-(\mathfrak{h}'-\mathfrak{h})\mathrm{R},$$

and on transference to the next surface

$$\mathfrak{h}_{m+1}=\mathfrak{h}'_m+t_m\sin\psi_m,$$

where $\mathfrak{h}$ and $\mathfrak{h}'$ are the perpendiculars to the ray from the vertex of the surface before and after refraction, and $\sin\psi_0'$ is the approximate value of the sine of the inclination after refraction given by the paraxial law

$$\sin\psi_0'=\sin\psi+\sin\phi'-\sin\phi.$$

It will be noted that the aberration terms are those involving $\mathfrak{h}'-\mathfrak{h}$, and the terms in the numerator of the expression for this quantity are the first order aberration terms. The magnitude of the denominator in comparison with its paraxial value 8 gives the ratio of the terms of higher orders to the first order values. Rays traced through by this method indicate directly where departures from the approximate expressions occur and what measures should be taken to secure the most satisfactory compromise or, where possible, to secure thorough correction. The only tables needed are a single-page, four-figure sheet giving the cosines from the sines. It will be noted that the differences between the cosines and unity occur as a small correcting factor in the aberrational expression, and the four figures give approximately a degree of accuracy comparable with that obtainable from six- or seven-figure logarithms using the ordinary trigonometrical method of ray tracing.

A consideration which will appeal strongly to the computer who is working continuously at ray tracing is the reduced mental demand the use of this method involves. It has been devised particularly for use where calculating machines are available, and it is only under these conditions that its special advantage will be realised.

T. S.

OPTICAL CONSTANTS, TABLE OF, FOR TYPICAL OPTICAL GLASSES. See "Glass," § (23).

OPTICAL GLASS

§ (1) PROPERTIES OF OPTICAL GLASS.—The material most generally used for the construction of the prismatic and lenticular portions of optical instruments is glass. This is available in many varieties, all specially manufactured for the purpose from the purest materials obtainable, and so treated that the finished product is as transparent as possible for the whole visible spectrum and the contiguous spectral regions, is homogeneous, free from internal strain and from certain other defects. In certain types of glass the entire removal of some of the minor defects has not been found possible. For example, dense baryta crowns almost invariably contain throughout the whole mass of the melting a number of bubbles. If these are small and not too numerous they are of little consequence in instruments in which the employment of these glasses is of special importance, and the glass-maker contrives to cut out the larger bubbles and any pieces in which the bubbles are too numerous before moulding the remainder into plates or discs for the use of the lens-maker. Again, certain glasses which in other respects possess properties of great value are very liable to tarnish or even to decompose after prolonged exposure to the atmosphere, and their use is practically confined to combinations in which they can be cemented between two lenses of more durable glass. A few very dense flint glasses begin to absorb light strongly before the violet end of the spectrum is reached, and in considerable thicknesses they are practically opaque to blue light. Their use is consequently confined to the construction of very small lenses, such as the lenses of microscope objectives of high power, where the short length of the path of light within the glass renders its strong absorption almost innocuous.

Apart from the presence of objectionable qualities such as those just mentioned, particular types of glass are selected solely for their purely optical properties. The optical computer could desire the liberty to select arbitrarily the two principal properties of the glass he would use, the dispersion which controls the chromatic aberration, and the refractive index for a selected colour which determines the spherical errors of the instrument. Until new glasses were manufactured at Jena by Schott u. Gen. the choice was limited to glasses of a linear series. If μ is used to denote refractive index, and a suffix is added to distinguish the spectral line to which it relates, the dispersion is measured by the magnitude of ν where

$$\mu_F - \mu_C = \frac{(\mu_D - 1)}{\nu}. \qquad (1)$$

The older glasses obey a relation of the form

$$\frac{\alpha}{\mu_D} + \frac{\beta}{\nu} = 1, \qquad (2)$$

where α and β are constants having the same values for all the glasses.

Consider now an object glass made up of a number of thin lenses in contact, each kind of glass satisfying the above relation. Let R be the difference in the curvatures of a typical component, and write κ for R $(\mu - 1)$, adding a suffix when μ carries one. From equation (1)

$$\Sigma\kappa_F - \Sigma\kappa_C = \Sigma\frac{\kappa_D}{\nu}, \qquad (3)$$

showing that the C and F rays will only be similarly refracted if

$$\Sigma\frac{\kappa_D}{\nu} = 0, \qquad (4)$$

which is the usual condition for achromatism in a telescope objective. Multiplying both sides of (2) by κ_D,

$$\alpha\Sigma\frac{\kappa_D}{\mu_D} + \beta\Sigma\frac{\kappa_D}{\nu} = \Sigma\kappa_D. \qquad (5)$$

It is ordinarily essential for the condition (4) to be closely satisfied, so that the term in β disappears from (5). The quantity on the right is the power of the system, and the coefficient of α is the quantity which determines the curvature of the field. It follows that with such glasses there is no power to vary the curvature of the field of a thin objective.

The realisation of the limitations implied by the relation (2), together with another property of optical glasses to be mentioned presently, induced a number of experimenters to search for glasses outside the region to which known types belonged. Among these Vernon Harcourt and Stokes may be mentioned. Their investigations achieved only a partial success and led to no results of practical utility. It turned out that they just missed success, largely through the want of such technical advice as a glass-maker could afford. Abbe, working in conjunction with Schott, an experienced glass manufacturer, succeeded

in overcoming the difficulties of making several novel varieties of glass in about two years. The results of their work were of such importance that the Jena glassworks were founded for the manufacture of the new glasses. A first list of forty-four types of glass was published in 1886, of which nineteen represented glasses of novel optical position. Other lists have been issued since at various dates, and taking all the lists together glasses have been listed in the regions indicated in *Fig.* 1, which shows a very different state of

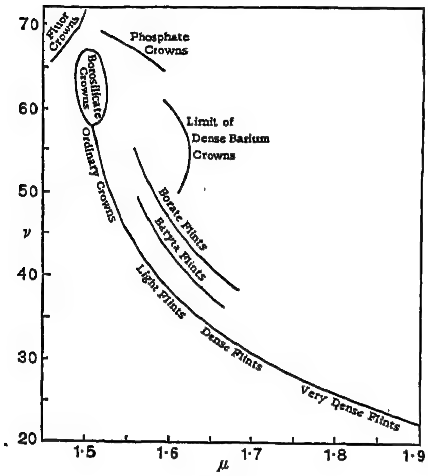


FIG. 1.

affairs from the hyperbolic line which represented the glasses obtainable earlier.

The most important of the new varieties have been:

Phosphate crowns	First introduced 1886
Barium phosphate crowns	
Boro-silicate crowns	
Dense barium crowns	
Borate flints	
Baryta light flints . .	First introduced 1888
Telescope crowns . . .	„ „ 1902
Telescope flints . . .	„ „ 1905
Fluor crowns	„ „ 1913

Of these glasses much the most remarkable in their properties were the phosphate glasses and borate flints of the first list. These have all been found to be unstable and have been withdrawn. In successive lists it will be noted that there has been a general tendency to continue to withdraw the more extreme glasses, and it may be inferred that stable glasses have not yet been made having extreme properties in desirable directions.

§ (2) IRRATIONALITY OF DISPERSION.—The production of glasses not constrained by the normal relation between refractive index and dispersion is the most notable permanent result of the work of the Jena glassworks. But though the importance of this achievement was not unappreciated, nevertheless it was not the prime motive which inspired the researches of Harcourt, Stokes, and Abbe. Much more attention was given in these investigations to the production of glasses with a considerable difference of dispersive power, but with a similar division of the dispersion for various parts of the spectrum. If two prisms are made from an ordinary crown and an ordinary flint so as to produce as closely as possible the same angular dispersion in the neighbourhood of the region of the spectrum to which the eye is most sensitive, it will be found that the crown prism will have the greater dispersion between this region and the red end of the spectrum, while the flint will produce the greater dispersion between this central region and the blue end of the spectrum. The consequence of this irrationality of dispersion is that a telescope objective, instead of bringing light of all visible wave-lengths sensibly to the same focus, has a minimum focal length for some wave-length which may be chosen by the designer, and for wave-lengths towards the ends of the spectrum the position of focus is so different that images in light of these colours will not be seen. Moreover, the longer the focal length the greater is the defect, so that it is particularly serious in all large refracting telescopes. The relatively advanced degree of development of astronomical instruments at an early stage in the evolution of optics led to early recognition of this fault, and thus it naturally called for the attention of those who sought to improve optical glass. The importance of the other direction in which freedom was desired, the removal of the restriction expressed by equation (2), was only recognised as of considerable importance with the development of photography.

From equations (1) and (3) it is evident that in a system for which C and F have a common focus, so that equation (4) is satisfied, the focus for some other line, say X, will coincide with C and F if all the glasses have the same value for

$$\frac{\mu_X - \mu_C}{\mu_F - \mu_C}.$$

A ready index of the suitability of two glasses for producing very perfect colour correction will thus be afforded by tabulating these ratios where X is made to agree in turn with a number of spectral lines. Such ratios are given in the lists issued by glass-makers, the usual ratios chosen being

$$\frac{\mu_D - \mu_{A'}}{\mu_F - \mu_C}, \quad \frac{\mu_F - \mu_D}{\mu_F - \mu_C}, \quad \frac{\mu_{G'} - \mu_F}{\mu_F - \mu_C}.$$

It is not easy, however, from an examination of these figures, which must be given to three decimal places to be of practical utility, to realise what has actually been achieved. An alternative method will therefore be adopted which will show the degree of uniformity clearly with the aid of the minimum number of figures. The method has the further merit of leading directly to a number of conclusions of considerable importance.

A rough examination of the figures of the glass lists will show that the change of distribution of the dispersion runs fairly uniformly with ν. We may, therefore, fit a formula of the type

$$\mu - \mu_D = (\mu_F - \mu_C)(a + b\nu)$$

to the glasses as a whole for the various spectral intervals, and tabulate the outstanding corrections which must be made to the calculated dispersions to equal those observed. It is preferable to apply this method to the actual differences of refractive index rather than to the calculated ratios, for the former are likely to be more accurate and the errors of the measurements are approximately constant in index differences. As a basis it is important to employ a list which includes many varieties of glass, both ordinary and new types, and the list should be critically examined for any internal evidence of unreliability in the figures. The various Jena glass lists, though not entirely free from error, will be found the most satisfactory for the purpose, and these lists have been utilised in preparing the following tables.

The formula for the refractive index of any given wave-length now becomes

$$\mu - \mu_D = a(\mu_F - \mu_C) + b(\mu_D - 1) + \epsilon \times 10^{-5}, \quad (6)$$

where a and b are functions of the wave-length only and ϵ is the divergence from normality. When the wave-length for the D line is substituted into the formulae for a and b they must both vanish. Moreover b has the same value for the C and F lines, and for these lines a has values differing by unity. It is found that the outstanding values of ϵ are small if the following values are adopted for these constants:

Line.	Wave-length.	a.	b.
A′	0·7682	– ·544	– ·00167
C	0·6563	– ·269	– ·00045
D	0·5893	0	0
F	0·4861	·731	– ·00045
G′	0·4341	1·404	– ·00223

It is important to notice the meaning to be attached to b in lenses composed of glasses having $\epsilon = 0$. By addition for the various glasses (6) gives

$$\Sigma\kappa = a(\Sigma\kappa_F - \Sigma\kappa_C) + (1 + b)\Sigma\kappa_D. \quad (7)$$

Thus in a system achromatised for C and F the power for any particular wave-length is $1 + b$ times that for the D line. Thus b measures the so-called secondary spectrum, and this is invariable for combinations of normal glasses forming a thin lens. It will be seen later that b assumes small positive values somewhat on the short wave-length side of the D line, so that the extreme variation of the focus is somewhat greater than the preceding figures suggest. By simple transformation the reference position may be chosen at any other part of the spectrum in place of the D line, and the precise type of chromatic correction may also be chosen arbitrarily, without altering the general character of the chromatic relations. The equation to the secondary spectrum can readily be found under any assigned conditions by expressing b as a function of the wave-length.

§ (3) Refractive Index and Wave-length.—Most attempts at fitting a formula to express the refractive index of glass as a function of the wave-length appear to have been based upon the study of a few glasses for which the indices have been observed for a large number of lines. Such a method is doubtless satisfactory for the particular glasses to which the formula is fitted, but it almost invariably happens that these formulae fail seriously when applied to normal glasses in general. The fact that a formula such as (6) expresses a relation which holds within narrow limits for such glasses indicates that a high degree of accuracy can be attained from a formula which involves only two distinct functions of the wave-length. To determine these functions let the term in $(\mu_D - 1)$ be eliminated from the equations, and to obtain a good distribution of the observed points let only $\mu_{A'}$, μ_C, μ_F, and $\mu_{G'}$ enter into the equation so obtained. The relation is evidently

$$\begin{aligned}(b_{G'} - b_F)(\mu_C - \mu_{A'}) - (b_C - b_{A'})(\mu_{G'} - \mu_F) \\ + \{(a_C - a_{A'})(b_{G'} - b_F) - (a_{G'} - a_F)(b_C - b_{A'})\} \\ (\mu_F - \mu_C) = 0,\end{aligned}$$

or, numerically,

$$178(\mu_C - \mu_{A'}) + 122(\mu_{G'} - \mu_F) = 131(\mu_F - \mu_C).$$

A shorter very approximate relation is

$$17(\mu_C - \mu_{A'}) + 11(\mu_{G'} - \mu_F) = 12(\mu_F - \mu_C).$$

Now the glasses have two quite distinct degrees of freedom, but all obey this relation. It follows that each function of the wave-

lengths must obey this relation separately. The great majority of the functions that may be proposed may be ruled out of court at once since this relation is not satisfied by them.

A class of functions which may be considered on account of their simplicity is that where the wavelength occurs to some power n. As n is given different values, the value assumed by

$$178\{(\cdot6563)^n-(\cdot7682)^n\}-131\{(\cdot4861)^n-(\cdot6563)^n\}+122\{(\cdot4341)^n-(\cdot4861)^n\}$$

or

$$122(\cdot4341)^n-253(\cdot4861)^n+309(\cdot6563)^n-178(\cdot7682)^n$$

may be plotted. It is found that the resulting curve (see *Fig.* 2) crosses the zero line for two values of n,[1] viz. $-0{\cdot}91$ and $-3{\cdot}40$. The corresponding formula for the refractive index is readily seen to be

$$\mu-\mu_D=\{\cdot226(\mu_F-\mu_C)+\cdot0062(\mu_D-1)\}\frac{\lambda^{-\cdot91}-\lambda_D^{-\cdot91}}{\lambda_F^{-\cdot91}-\lambda_C^{-\cdot91}}+\{\cdot774(\mu_F-\mu_C)-\cdot0062(\mu_D-1)\}\frac{\lambda^{-3\cdot4}-\lambda_D^{-3\cdot4}}{\lambda_F^{-3\cdot4}-\lambda_C^{-3\cdot4}},$$

and from this the equation to the secondary spectrum for any type of correction is deducible without difficulty. When C and F are brought together the minimum focus is found at wave-length $\lambda=0{\cdot}5558$ with $b=+0{\cdot}00008$, and when D and G′ are united as for photographic purposes, b takes the values

$b_{A'}=-\cdot00253$, $b_F=+\cdot00071$,

$b_C=-\cdot00088$, $b_{G'}=0$,

$b_D=0$, $b_{max.}=+\cdot00071$ at $\lambda=\cdot4972\mu$.

The formula found above shows (*Fig.* 3) that of those hitherto proposed Conrady's[2] second form

$$\mu=\mu_0+a\lambda^{-1}+b\lambda^{-\frac{7}{2}}$$

is that of most general utility when only two functions are employed. If the Cauchy form

$$\mu=\mu_0+a\lambda^{-2}+b\lambda^{-4}$$

were assumed, the relation

$$494(\mu_C-\mu_{A'})+245(\mu_{G'}-\mu_F)=300(\mu_F-\mu_C)$$

would be obeyed; it may be shown that no glass obeys this law, and the formula will therefore be unreliable when applied to the dispersion of glass. Other formulae may readily be found which will obey the actual law as accurately as that already determined, but in view of the simplicity of this result and the present degree of accuracy in refractometric observations there is little gain to be expected therefrom.

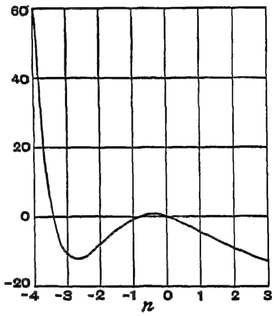


FIG. 2.

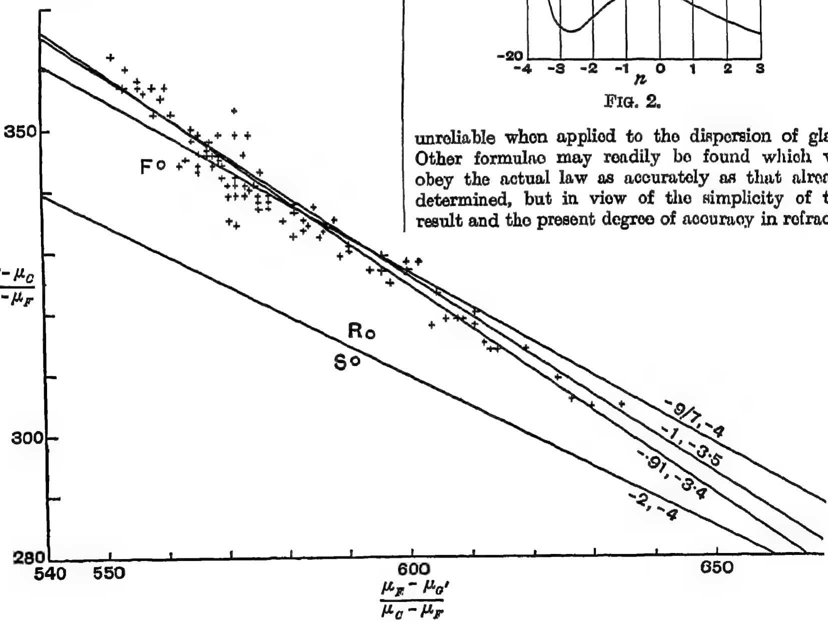


FIG. 3.

F = Fluorite. R = Rock salt. S = Sylvite.

[1] The solution $n=0$ is, of course, valueless.

[2] *Monthly Notices, R.A.S.* lxiv. 460.

§ (4) Table of Properties.—In the following tables the glasses, as usual, are arranged in the order of their ν values. The type number is given first—O. indicates that the ordinary procedure is followed in manufacture, S. that the glass is made in special pots on a small scale only. As the history of the glass is of interest, the date of introduction is given, and if the glass has since been withdrawn the date of withdrawal also.

For brevity the lists are denoted by letters, the dates being as follows:

1886	A	1902	D
1888	B	1905	E
1892	C	1913	F

Where two letters appear, the earlier indicates the list in which the glass is first mentioned, the latter the first list from which the glass is omitted. In the descriptions of the glasses the following contractions are used:

Ba.	Barium	h.	High
B.	Borate	l.	Low
BS.	Boro-silicate	L.	Light
Cr.	Crown	M.	Medium
d.	Dense	O.	Ordinary
D.	Densest	P.	Phosphate
Δ.	Dispersion	S.	Silicate
E.	Extra	T.	Telescope
F.	Fluor	UV.	Ultra-violet
Fl.	Flint	Z.	Zinc
G.	Glass		

Type.	Description.	Introduced and withdrawn.	ν.	$(\mu_D - 1)10^4$.	$(\mu_F - \mu_C)10^5$.	$e_{A'}$.	$e_C = e_F$.	$e_{G'}$.
O. 225	L. P. Cr.	A : D	70·0	5159	737	+ 2	− 1	− 2
O. 6781	F. Cr.	F	69·9	4933	706	− 1	..	+ 3
O. 6500	F. Cr.	F	67·1	4711	702	− 2	− 1	− 1
S. 40	M. P. Cr.	A : D	66·9	5590	835	+ 2	+ 2	+ 5
O. 82	BS. Cr.	D : E	66·5	4944	743	− 9	− 2	− 2
O. 3296	BS. Cr.	E : F	66·5	4991	750	− 8	− 2	− 1
O. 3258	Cr. l. μ.	D : F	65·9	4782	726	− 11	− 2	− 6
O. 2188	BS. Cr.	D	65·9	5013	760	− 1	..	+ 2
O. 7185	F. Cr.	F	65·6	4637	707	− 2	..	..
O. 3467	Cr. l. μ.	E : F	65·6	4649	708	− 8	− 3	..
S. 30	d. Ba. P. Cr.	A : D	65·2	5760	884	+ 7	+ 2	+ 7
O. 802	BS. Cr.	C	64·9	4967	765	− 5	− 3	− 6
√UV. 3199	UV. Cr.	E	64·4	5035	781	− 5	− 2	− 6
S. 15	D. Ba. P. Cr.	A : D	64·1	5906	922	+ 9	+ 1	+ 6
O. 144	BS. Cr.	A	64·0	5100	797	..	− 1	− 2
O. 3832	BS. Cr.	E	64·0	5163	806	− 3	..	− 2
O. 3848	BS. Cr. h. μ.	E	63·5	5199	818	− 5	− 2	− 1
O. 599	BS. Cr.	B	62·3	5069	813	− 2	− 2	− 2
O. 3512	BS. Cr. h. μ.	E	62·3	5301	851	− 5	+ 1	− 2
O. 57	L. S. Cr.	A	61·8	5086	823	+ 3	− 1	..
O. 2388	T. Cr.	D : E	61·7	5254	852	+ 2	+ 3	+ 7
O. 3390	BS. Cr.	E	61·2	5170	844	+ 1	+ 1	+ 3
O. 6367	BS. Cr.	F	61·0	5088	835	− 6	− 1	− 5
O. 3655	T. Cr.	E : F	61·0	5285	866	+ 2	+ 1	+ 5
O. 40	S. Cr.	A : D	60·9	5166	849	+ 3	− 1	− 2
O. 2122	D. Ba. Cr. h. ν.	D	60·8	5899	970	+ 5	..	− 1
O. 337	S. Cr.	B	60·7	5144	847	..	..	+ 2
O. 4817	S. Cr.	F	60·6	5028	830	+ 1	..	..
O. 374	S. Cr.	B : D	60·5	5109	844	− 3	− 1	+ 1
O. 6223	S. Cr.	F	60·5	5118	846	..	..	− 1
O. 3453	S. Cr.	E	60·4	5191	860	+ 4	..	− 1
O. 546	Z. S. Cr.	B	60·2	5170	859	− 1	..	− 1
O. 60	S. Cr.	A	60·2	5179	860	+ 1	..	..
O. 138	S. Cr. h. μ.	A	60·2	5258	872	+ 2	..	+ 1
S. 52	L. B. Cr.	A : D	60·0	5047	840	− 19	− 4	− 14
O. 4125	S. Cr.	F	60·0	5223	871	+ 1	..	+ 1
O. 6634	S. Cr.	F	59·7	5127	858	− 1	..	..
O. 567	S. Cr.	B	59·7	5134	859	− 1	..	+ 1
O. 20	S. Cr. l. μ.	A : D	59·6	5019	842	− 1	− 1	..
O. 2118	Cr. l. μ.	D	59·4	5095	858	− 5	..	+ 4
O. 227	Ba. S. Cr.	A	59·4	5309	900	+ 3	− 1	− 3
O. 3712	D. Ba. Cr.	E	59·3	6053	1021	+ 1	..	..
O. 610	Cr. l. μ.	B : D	59·0	5063	858	− 1	− 2	− 1
O. 203	O. S. Cr.	A	59·0	5175	877	..	− 2	− 1
O. 2071	D. Ba. Cr.	D	58·8	6008	1037	+ 1	− 1	..
O. 598	S. Cr.	B : D	58·6	5152	879	+ 2	..	− 1
O. 512	S. Cr.	B : D	58·6	5192	886	+ 1	+ 1	+ 1

Type.	Description.	Introduced and withdrawn.	ν.	$(\mu_D-1)10^4$.	$(\mu_F-\mu_C)10^5$.	$\epsilon_{A'}$.	$\epsilon_C=\epsilon_F$.	$\epsilon_{G'}$.
O. 2164	Cr. L. μ.	D	58·4	5102	873	+ 1	+ 1	+ 4
O. 13	Pot. S. Cr.	A : D	58·0	5228	901	+ 5	+ 2	+ 4
O. 15	Z. S. Cr.	A	58·0	5308	915	− 1	− 1	− 2
O. 211	D. Ba. S. Cr.	A	57·5	5726	995	+ 7	. .	+ 1
O. 709	Z. Soda Cr.	C : D	57·3	5128	894	− 3	. .	− 3
O. 3376	Soft S. Cr.	E	57·3	5189	904	. .	+ 1	. .
O. 3551	Z. S. Cr.	E	57·2	5127	897	− 3	− 1	. .
O. 153	S. Cr.	A : D	57·2	5160	904	+ 2	− 1	− 2
O. 1209	D. Ba. Cr.	C	57·2	6112	1068	+ 3	. .	. .
O. 5970	D. Ba. Cr.	F	56·9	6207	1092	+ 4	. .	− 2
O. 114	Soft S. Cr.	A	56·6	5151	910	+ 4	. .	. .
O. 197	B. S. G.	A : D	56·5	5250	929	− 6	− 1	− 2
O. 1615	D. Ba. Cr.	D	56·4	6080	1078	+ 3	. .	. .
O. 2994	D. Ba. Cr.	D	56·4	6130	1087	+11	+ 1	− 1
O. 7550	Ba. L. Fl.	F	56·0	5694	1017	+ 1	. .	− 1
O. 3961	D. Ba. Cr.	F	55·8	6119	1097	+10	+ 1	− 1
O. 7336	Ba. L. Fl.	F	55·7	5682	1020	+ 3	. .	− 2
O. 3775	D. Ba. Cr.	E : F	55·7	6116	1098	+ 4	+ 2	+ 4
UV. 3248	UV. Fl.	E	55·4	5332	964	+ 2	− 1	− 2
O. 463	Ba. L. Fl.	B	55·4	5646	1020	+ 1	. .	. .
O. 202	D. Ba. S. Cr.	A : D	55·3	6040	1092	+ 5	. .	− 1
S. 35	B. Fl.	A : D	55·2	5503	996	−20	− 4	− 16
O. 5878	D. Ba. Cr.	F	54·9	6139	1117	+ 3	+ 1	. .
O. 608	Cr. h. Δ.	B	54·6	5149	943	+ 4	. .	. .
O. 3419	T. Fl. h. ν.	E : F	54·6	5154	944	− 9	− 2	− 4
O. 4679	D. Ba. Cr.	F	54·1	6167	1140	+ 4	+ 1	. .
O. 252	B. Fl.	A : D	53·8	5521	1026	−17	− 3	− 13
O. 722	Ba. L. Fl.	C	53·8	5797	1078	+ 2	− 1	− 2
O. 5799	D. Ba. Cr.	F	53·2	6223	1170	+ 4	+ 1	− 1
O. 846	Ba. L. Fl.	C	53·0	5525	1042	+ 2	− 1	− 2
O. 602	Ba. L. Fl.	B	53·0	5676	1072	+ 3	+ 1	− 2
O. 2001	T. Fl.	D : E	51·8	5211	1007	− 4	− 3	− 11
O. 3439	T. Fl.	E	51·6	5286	1025	− 8	− 2	− 7
O. 381	Cr. h. Δ.	B	51·3	5262	1026	+ 2	+ 1	. .
O. 152	S. G.	A	51·2	5368	1049	+ 1	. .	. .
O. 583	Ba. L. Fl.	B	51·2	5688	1110	+ 3	. .	− 2
O. 3422	T. Fl.	E : F	50·9	5257	1032	−10	− 3	− 8
S. 8	B. Fl.	A : D	50·8	5736	1120	−18	− 4	− 18
O. 543	Ba. L. Fl.	B	50·6	5637	1115	+ 2	. .	. .
O. 527	Ba. L. Fl.	B	50·4	5718	1133	+ 6	. .	. .
O. 3338	T. Fl. l. ν.	E	49·9	5483	1099	− 5	− 2	− 7
O. 164	BS. Fl.	A : F	49·4	5503	1114	−12	− 4	− 11
O. 2015	D. Ba. Cr. h. Δ.	D	49·4	6041	1222	+ 3	+ 1	− 2
O. 575	Ba. L. Fl.	B	49·3	5682	1151	+ 3	+ 1	. .
O. 214	S. G.	A : D	48·7	5366	1102	− 1	. .	− 3
O. 522	Ba. L. Fl.	B	48·2	5554	1153	+ 2	+ 1	+ 1
O. 7821	BS. Fl.	F	48·1	5703	1185	− 4	− 1	− 4
O. 726	E. L. Fl.	C	47·3	5398	1142	. .	. .	− 4
O. 6241	O. L. Fl.	F	47·0	5595	1192	− 2	. .	− 2
O. 161	BS. Fl.	A : F	46·7	5676	1216	− 6	− 3	− 12
O. 578	Ba. L. Fl.	B	46·4	5825	1255	+ 3	. .	− 2
O. 378	E. L. Fl.	C	45·9	5473	1193	+ 1	. .	− 1
O. 364	BS. Fl.	D : F	45·9	5753	1254	− 9	− 3	− 9
O. 6296	O. L. Fl.	F	45·4	5584	1230	− 1	. .	− 2
S. 7	B. Fl.	A : D	44·3	6086	1375	−14	− 4	− 19
O. 1266	Ba. L. Fl.	D	43·8	6042	1381	+ 1	. .	− 1
O. 154	O. L. Fl.	A	43·0	5710	1327	− 2	− 1	. .
O. 376	O. L. Fl.	B	42·9	5660	1319	− 2	. .	. .
O. 230	S. Fl. h. μ.	A : D	42·5	6014	1415	+ 2	+ 2	− 1
O. 276	O. L. Fl.	D	42·2	5800	1373	− 2	− 1	+ 6
O. 569	O. L. Fl.	B	41·4	5738	1385	− 4	. .	+ 1
O. 340	O. L. Fl.	B	41·4	5774	1396	− 1	. .	. .
O. 184	O. L. Fl.	A	41·0	5900	1438	− 1	− 3	− 4
S. 17	d. B. Fl.	A : D	40·6	6467	1501	−16	− 6	− 25
O. 748	Ba. Fl.	C	39·1	6235	1599	+ 9	+ 1	+ 1

Type.	Description.	Introduced and withdrawn.	ν.	$(\mu_D-1)10^4$.	$(\mu_F-\mu_C)10^5$.	$\epsilon_{A'}$.	$\epsilon_C=\epsilon_F$.	$\epsilon_{G'}$.
O. 318	O. L. Fl.	B	38·3	6031	1575	− 2	..	− 1
S. 10	d. B. Fl.	A : D	38·0	6797	1787	−11	− 5	− 24
O. 118	O. Fl.	A	36·9	6129	1660	− 1	− 2	− 2
O. 167	O. Fl.	A	36·5	6169	1691	− 3	− 2	− 2
O. 3269	D. Ba. Fl.	D	36·3	6570	1809	+ 1	+ 2	+ 8
O. 103	O. Fl.	A	36·2	6202	1709	− 1	− 1	..
O. 3863	O. Fl.	E : F	35·9	6223	1731	+ 3	+ 2	+ 6
O. 93	O. Fl.	A	35·8	6245	1743	..	− 3	− 2
O. 6131	O. L. Fl.	F	35·7	6246	1752	+ 2	+ 1	+ 3
O. 919	O. Fl.	D	35·7	6315	1770	+ 5	+ 1	+ 7
O. 266	O. Fl.	B : D	35·4	6287	1775	− 1	+ 1	+ 4
O. 335	d. Fl.	B	34·8	6372	1831	+ 3	− 2	+ 3
O. 102	d. Fl.	A	33·8	6482	1919	..	− 2	+ 3
O. 192	d. Fl.	A	32·0	6734	2104	+ 2	− 1	+ 5
O. 41	d. Fl.	A	29·5	7174	2434	+ 5	+ 2	+ 13
O. 113	d. Fl.	A	28·4	7371	2600	+11	+ 3	+ 16
O. 165	d. Fl.	A	27·5	7541	2743	+11	+ 3	+ 21
O. 198	D. Fl.	A	26·5	7782	2941	+11	+ 5	+ 32
S. 228	D. Fl.	D : E	21·7	9044	4174	+28	+12	+ 90
S. 57	D. Fl.	A : D	19·7	9626	4882	+70	+22	+159

It should be noted in considering the values of the ϵ's that the dispersions are supposed to be correct to ± 2 units in the fifth decimal place. An error of this amount in the measured value of $\mu_F-\mu_C$ would introduce an error of 3 in the calculated value of $\mu_{G'}$, and as there might be an error in the measurement of $\mu_{G'}-\mu_D$ of 2, it is possible for an error of 5 units to appear in the last column without any actual deviation in the glass from normal behaviour. The figures suggest that errors only reach such a magnitude in one or two isolated cases, a check being obtained by comparing the behaviour of any glass with the normal properties of other glasses of the same type, and particularly with those of neighbouring optical position. If glasses of higher index than 1·7 be excepted in addition to those that have been withdrawn on account of their instability, there remain very few types whose departure from the normal distribution of the dispersion exceeds the experimental errors of the measurements. It may thus be said that, apart from unstable types, practically no success has been achieved in the manufacture of glass suitable for the elimination of secondary spectrum. For all stable glasses the dispersion formula already mentioned is sufficiently accurate, but for special glasses such as borate flints an additional term in the formula is necessary.

It is not unimportant to note what values are required in the ϵ's for two glasses to be used for a thin achromatic lens to enable the secondary spectrum to be eliminated. Let the outstanding ϵ's be divided into two portions, the one proportional to the medium dispersions of their glasses, the rest proportional to the refracting power but of opposite sign in the two glasses. The former portion may be neglected, since it simply gives a corrected value to a in formula (6). The indices may now be written in the form

$$\mu-1=(\mu_D-1)\left\{1+\frac{a}{\nu}+b+c\frac{\nu-\nu'}{\nu+\nu'}\right\},$$

$$\mu'-1=(\mu_D'-1)\left\{1+\frac{a}{\nu'}+b-c\frac{\nu-\nu'}{\nu+\nu'}\right\},$$

and therefore

$$\Sigma\kappa=(1+b+c)\Sigma\kappa_D$$

for an achromatic combination. There will be no secondary spectrum if $b+c=0$ for all colours. Since b is negative for the reference lines other than D, the ϵ's must be positive for the crown lens and negative for the flint. Taking 1·55 as a mean value for both indices, the ϵ's of each lens must depart from the normal to the following extents for the ν values shown to bring the focus for the lines indicated to the same position as that for the D line:

ν.	ν'.	$\epsilon_{A'}$.	$\epsilon_C=\epsilon_F$.	$\epsilon_{G'}$.
70	30	37	10	50
65	35	28	7	37
60	40	18	5	25
55	45	9	2	12

Since there are no crown glasses available which are above normal even as much as is indicated in the last rows of figures, the greater part of the correction will have to be borne by the low values of the flint glass. These figures indicate clearly that a ν difference of approximately 10 was the extent of the achievement reached for objectives without secondary spectrum even with unstable glasses,

The following table gives the mean values of results obtained by various observers (Landolt-Börnstein Tables, made at the Reichsanstalt (*Z. f. Instrumentenkunde*, 1920, xl. 94); they are for a temperature of 20° C. and refer rock salt and sylvite are for a temperature of 18° C. For particulars with regard to the positions of infra-red

TABLE I.—REFRACTIVE INDICES

O = ordinary ray.

	Calcite ($CaCO_3$).			Fluorite (CaF_2).			
	Wave-length.	O.	E.	Wave-length.		Wave-length.	
	μμ.			μμ.		μμ.	
Ultra-violet.	Al 199·0	..	1·57796	Al 185·4	1·51006	..	..
	Au 200·1	1·90284	1·57649	Al 193·5	1·50137	..	..
	Au 204·5	1·88242	1·57081	Al 199·0	1·49628	..	..
	Au 208·2	1·86733	1·56640	Zn 202·5	1·49325	..	..
	Zn 210·0	1·86081	..	Zn 206·2	1·49033	..	..
	Cd 214·4	1·84584	1·55990	Zn 210·0	1·48753	..	..
	Cd 219·5	1·83082	1·55511	Cd 214·4	1·48461	..	..
	Cd 226·5	1·81305	1·54924	Cd 219·5	1·48154	..	..
	Cd 231·3	1·80245	1·54554	Cd 226·5	1·47758	..	..
	Ag 244·6	1·77966	1·53731	Cd 231·3	1·47522	..	..
	Cd 257·3	1·76055	1·53017	Ag 244·6	1·46965	..	..
	Cd 274·9	1·74152	1·52272	Cd 257·3	1·46481	..	..
	Sn 303·4	1·71959	1·51365	Cd 274·9	1·45967	..	..
	Zn 330·3	1·70515	1·50746	Sn 303·4	1·45338	..	..
	Cd 361·1	1·69317	1·50224	Zn 330·3	1·44907	..	..
	..	..	..	Cd 361·1	1·44534	..	..
	μμ.			μμ.		μμ.	
Visible.	Al 396·2	1·68330	1·49777	Al 396·2	1·44219	Hg 404·7	1·44151
	H 434·1	1·67551	1·49428	H 434·1	1·43962	Hg 435·8	1·43949
	H 486·1	1·66785	1·49073	H 486·1	1·43707	He 447·2	1·43887
	Cd 508·6	1·66527	1·48959	Cd 508·6	1·43620	Zn 472·2	1·43764
	Fe 527·0	1·66341	1·48871	Fe 527·0	1·43557	H 486·1	1·43704
	Cd 533·8	1·66276	1·48842	Cd 533·8	1·43536	He 501·8	1·43643
	Pb 560·9	1·66046	1·48735	Hg 546·1	1·43497	Cd 508·6	1·43617
	Na 589·3	1·65837	1·48641	Pb 560·9	1·43457	Hg 546·1	1·43496
	H 656·3	1·65440	1·48458	Na 589·3	1·43385	Hg 578·0	1·43410
	Li 670·8	1·65369	1·48431	H 656·3	1·43251	Na 589·3	1·43383
	He 706·5	1·65207	1·48353	Li 670·8	1·43226	Zn 636·2	1·43284
	K 768·2	1·64974	1·48258	He 706·5	1·43171	Cd 643·9	1·43271
	Rb 795·0	1·64886	1·48216	K 768·2	1·43094	H 656·3	1·43248
	..	..	..	Rb 795·0	1·43064	He 706·5	1·43169
	μ.			μ.		μ.	
Infra-red.	0·8325	1·64772	1·48176	0·8840	1·42981	0·7707	1·43088
	0·9047	1·64578	1·48098	1·1786	1·42788	0·8191	1·43037
	0·9914	1·64380	1·48022	1·3756	1·42690	0·9610	1·42919
	1·0973	1·64167	1·47948	1·5715	1·42596	1·0922	1·42835
	1·2288	1·63926	1·47870	1·7680	1·42505	1·1560	1·42799
	1·3070	1·63789	1·47831	1·9153	1·42434	1·1786	1·42787
	1·3958	1·63637	1·47789	2·0626	1·42358	1·4416	1·42658
	1·4972	1·63457	1·47744	2·3573	1·42204	1·6382	1·42565
	1·6087	1·63261	..	2·6519	1·42017	1·7340	1·42520
	1·6146	..	1·47695	2·9466	1·41824	1·7679	1·42504
	1·7487	..	1·47638	3·2413	1·41611	2·0343	1·42373
	1·7614	1·62974	..	3·5359	1·41377	2·1843	1·42293
	1·9085	..	1·47573	3·8306	1·41119	2·3121	1·42222
	1·9457	1·62602	..	4·1252	1·40854	2·3572	1·42197
	2·0531	1·62372	..	4·7146	1·40237	2·5450	1·42086
	2·0998	..	1·47492	5·3036	1·39528	2·5754	1·42068
	2·1719	1·62099	..	5·8932	1·38717	..	..
	2·3243	..	1·47392	6·4825	1·37817	..	..
	..	..	..	7·6612	1·35680	..	..
	..	..	..	8·8398	1·33079	..	..
	..	..	..	9·4291	1·31612	..	..

4th ed., 1912, 969-973). The second column of values for fluorite are the results of recent measurements to air at 20° C., 760 mm. pressure, and 9 mm. absolute humidity. The other values for fluorite and those for absorption bands and the limits of transmission reference should be made to the Landolt-Börnstein Tables.

OF TRANSPARENT CRYSTALS

E = extraordinary ray.

Quartz (SiO_2).				Rock Salt (NaCl).		Sylvite (KCl).	
Wave-length.	O.	E.	Fused.	Wave-length.		Wave-length.	
μμ.				μμ.		μμ.	
Al 185·4	1·67580	1·68997	1·57464	Al 185·4	1·89332	Al 185·4	1·82704
Al 193·5	1·65996	1·67343	1·56003	Al 193·5	1·82809	Au 197	1·73114
Al 199·0	1·65090	1·66396	1·55201	Al 199·0	1·79580	Al 199·0	1·72432
Zn 206·2	..	..	1·54271	Au 200·1	1·79016	Au 200·1	1·71864
Cd 214·4	1·63041	1·64265	1·53392	Au 204·5	1·76948	Au 204·5	1·69811
Cd 219·5	1·62496	1·63701	1·52910	Au 208·2	1·75413	Au 208·2	1·68302
Cd 226·5	..	..	1·52305	Cd 214·4	1·73220	Cd 214·4	1·66182
Cd 231·3	1·61400	1·62560	1·51937	Cd 219·5	1·71710	Cd 219·5	1·64739
Ag 244·6	..	..	1·51096	Cd 226·5	1·69906	Cd 224·0	1·63606
Cd 257·3	1·59624	1·60713	1·50371	Cd 231·3	1·68843	Cd 231·3	1·62037
Cd 274·9	1·58752	1·59812	1·49613	Cd 257·3	1·64611	Au 242·8	1·60041
Sn 303·4	..	..	1·48688	Cd 274·9	1·62692	Cd 257·3	1·58119
Zn 330·3	..	..	1·48061	Cd 298·1	1·61226	Cd 274·9	1·56380
Cd 340·4	1·56743	1·57739	1·47877	Cd 340·4	1·58620	Al 308·2	1·54130
Cd 361·1	1·56347	1·57322	1·47511	Cd 361·1	1·57861	Cd 340·4	1·52720
..	..	..	..	..	..	Al 358·7	1·52109
μμ.				μμ.		μμ.	
Al 394·4	1·55846	1·56805	..	Cd 441·4	1·55962	Al 394·4	1·51213
Al 396·2	..	..	1·47054	H 486·1	1·55340	H 434·1	1·50493
H 434·1	1·55397	1·56340	1·46688	Fe 527·0	1·54915	H 486·1	1·49835
H 486·1	1·54967	1·55897	1·46317	Hg 546·1	1·54745	Cd 508·6	1·49610
Cd 508·6	1·54822	1·55746	1·46190	Pb 560·9	1·54629	Hg 546·1	1 49313
Fe 527·0	..	..	1·46099	Na 589·3	1·54432	Pb 560·7	1·49212
Cd 533·8	1·54680	1·55599	1·46067	H 656·3	1·54068	Na 589·3	1·49038
Pb 560·9	..	..	1·45951	Li 670·8	1·54002	H 656·3	1·48721
Na 589·3	1·54424	1·55334	1·45846	He 706·5	1·53863	Li 670·8	1·48663
H 656·3	1·54189	1·55091	1·45640	K 768·2	1·53666	K 768·2	1·48374
He 706·5	1·54050	1·54949	1·45516	..	..	..	..
K 768·2	1·53904	1·54797	1·45389	..	..	..	..
Rb 795·0	1·53851	1·54742	1·45340	..	..	..	..
..	..	..	..	..	..	..	..
μ.				μ.		μ.	
0·8007	1·53834	1·54725	..	0·8840	1·53401	0·8840	1·48132
0·8671	1·53712	1·54598	..	0·9822	1·53245	0·9822	1·47998
0·9460	1·53583	1·54404	..	1·1786	1·53037	1·1786	1·47821
1·0417	1·53442	1·54317	..	1·5551	1·52821	1·584	1·4765
1·1592	1·53283	1·54152	..	1·7680	1·52744	1·7680	1·47579
1·3070	1·53090	1·53951	..	2·3573	1·52585	2·3573	1·47465
1·3685	1·53011	1·53869	..	3·5359	1·52317	3·5359	1·47295
1·4219	1·52942	1·53796	..	4·1252	1·52164	4·125	1·4721
1·4972	1·52842	1·53602	..	5·0092	1·51898	5·3039	1·46991
1·6087	1·52687	1·53529	..	6·4825	1·51355	6·482	1·4678
1·6815	1·52583	1·53422	..	7·6611	1·50832	7·080	1·4660
1·7614	1·52468	1·53301	..	8·8398	1·50204	7·661	1·4645
1·9457	1·52184	1·53004	..	10·0184	1·49472	8·8398	1·46076
2·1719	1·51799	1·52609	..	12·9650	1·47172	10·0184	1·45662
2·59	1·5101	..	..	14·1436	1·46055	12·965	1·44336
3·03	1·4987	..	..	15·3223	1·44749	14·144	1·43712
3·40	1·4879	..	..	15·9116	1·44103	15·912	1·42607
3·80	1·4740	..	..	17·93	1·4151	17·68	1·41393
4·09	1·4620	..	..	20·57	1·3737	20·60	1·3882
56	2·18	..	..	22·3	1·3405	22·50	1·3692
..	..	..	..	..	..	..	..

and that appreciable reduction with present glasses can only be obtained by the use of abnormally steep curves. They further indicate the character of the chromatic focal relation to be expected in objectives corrected for secondary spectrum. In the outer parts of the spectrum there will be a tendency to approximate more closely to the normal type. Thus there is a general flattening of the focal curve, with four points in which the selected plane is crossed, not three as is usually stated. objective that can be made. The alternative is to have recourse to lenses of more complex structure in which an essential part in the correction is played by the separations of the components. That correction may be achieved in this way is readily seen. For consider a system built up of a number of thin achromatic lenses of ordinary glasses, separated from one another. Each of these produces a similar secondary spectrum, that is to say they behave as lenses all of the same kind of glass with

Table II.—Refractive Indices of Specimens of Optical Glass

	Wave-length.	Type of Glass.							
		S. 204.	O. 1151.	S. 179.	O. 451.	O. 1442.	O. 469.	O. 500.	S. 163.
	$\mu\mu$.								
Ultra-violet.	276·3	1·56027	..	..	..	..	..	..	..
	283·7	1·55648	..	..	..	..	..	..	..
	288·0	1·55437	..	..	..	..	..	..	..
	298·0	1·55005	1·57093	..	1·65397	..	..	..	..
	308·1	1·54625	1·56558	..	1·64453	..	..	..	..
	313·3	1·54444	1·56307	..	1·64024	1·65254	..	..	..
	326·1	1·54046	1·55770	1·59138	1·63134	1·64754	1·73245	..	..
	340·4	1·53660	1·55262	1·58776	1·62320	1·64271	1·71968	1·85487	..
	346·6	1·53509	1·55068	1·58632	1·62008	1·64077	1·71485	1·84731	..
	361·1	1·53195	1·54664	1·58330	1·61389	1·63683	1·70536	1·83263	..
	$\mu\mu$.								
Visible.	434·1	1·52092	1·53312	1·57273	1·59355	1·62320	1·67561	1·78800	1·94493
	467·8	1·51769	1·52903	1·56949	1·58772	1·61891	1·66725	1·77609	..
	480·0	1·51662	1·52782	1·56847	1·58594	1·61770	1·66482	1·77256	..
	486·1	1·51610	1·52715	1·56794	1·58515	1·61706	1·66367	1·77091	1·91890
	508·6	1·51447	1·52525	1·56643	1·58247	1·61504	1·65979	1·76539	..
	534·9	1·51287	1·52327	1·56476	1·57973	1·61292	1·65601	1·75995	1·90262
	589·3	1·51007	1·52002	1·56207	1·57524	1·60956	1·64985	1·75130	1·88995
	656·3	1·50742	1·51712	1·55957	1·57119	1·60644	1·64440	1·74368	1·87893
	768·2	1·50426	1·51368	1·55651	1·56669	1·60277	1·63820	1·73530	1·86702
	μ.								
Infra-red.	0·8	1·5044	1·5131	1·5555	1·5659	..	1·6373	1·7330	1·8650
	1·0	1·5009	1·5096	1·5522	1·5615	..	1·6315	1·7264	1·8541
	1·2	1·4979	1·5069	1·5497	1·5585	..	1·6277	1·7215	1·8481
	1·4	1·4950	1·5048	1·5476	1·5559	..	1·6246	1·7180	1·8433
	1·6	1·4919	1·5024	1·5452	1·5535	..	1·6217	1·7151	1·8396
	1·8	1·4884	1·4999	1·5424	1·5512	..	1·6193	1·7127	1·8364
	2·0	1·4845	1·4973	1·5390	1·5487	..	1·6171	1·7104	1·8334
	2·2	..	..	..	1·5463	..	1·6150	1·7082	1·8310
	2·4	..	..	..	1·5440	..	1·6131	..	1·8286

The failure of the attempt to manufacture stable glasses which will enable the secondary spectrum to be eliminated leads to the conclusion that, apart from new discoveries, two methods remain open for the construction of apochromatic objectives. The one consists in the employment of natural crystals of the cubic system for at least one component of the lens. Fluorite has been generally employed, and it is easy to see from the values of its dispersion given elsewhere that by its aid such objectives can be built. This method imposes severe limitations on the size of very small dispersion. If now the components are given the separations which would enable simple lenses all of the same glass and of the right powers to be chromatically corrected to the first order, it is evident that the condition for the removal of the secondary spectrum with the achromatic lenses is satisfied. That systems using only one glass may be chromatically corrected is well known, but they are of limited utility. A discussion of the procedure to be followed in applying either of these methods lies outside the scope of this article.

TABLE III.—EFFECT OF TEMPERATURE ON THE REFRACTIVE INDICES OF TRANSPARENT CRYSTALS

$\delta\mu$=alteration in absolute refractive index for sodium light per degree C. increase in temperature. O=ordinary ray. E=extraordinary ray.

Calcite.

Observer.	Mean Temp. °C.	$\delta\mu\times10^5$ O.	$\delta\mu\times10^5$ E.
Müller	5·7	+0·072	. .
Fizeau	42·5	+0·072	+1·103
Reed	57·1	+0·078	+1·094
Vogel	59·5	+0·089	+1·024
Micheli	61·5	+0·121	+1·106

Fluorite.

Observer.	Mean Temp. °C.	$\delta\mu\times10^5$.
Dufet	27	−1·34
Fizeau (Dufet)	33·5	−1·11
Baille	56·5	−1·122
Stefan	57·5	−1·24
Reed	58·8	−1·196
Pulfrich	60·5	−1·206
Micheli	61·25	−1·193

Quartz.

Observer.	Mean Temp. °C.	$\delta\mu\times10^5$ O.	$\delta\mu\times10^5$ E.
Müller	7	−0·589	. .
Fizeau (Dufet)	30	−0·598	−0·709
Dufet	50	−0·627	−0·741
Pulfrich	59·6	−0·638	−0·754
Reed	61·2	−0·607	−0·766
Micheli	61·4	−0·650	−0·754

Rock Salt.

Observer.	Mean Temp. °C.	$\delta\mu\times10^5$.
Lagerborg	52·5	−3·53
Stefan	56·5	−3·73
Baille	57·8	−3·70
Pulfrich	58·8	−3·739
Micheli	61·8	−3·733

Sylvite.

Observer.	Mean Temp. °C.	$\delta\mu\times10^5$.
Stefan	57·5	−3·46
Pulfrich	59·5	−3·641

Values of $\delta\mu$ for amorphous quartz for different wave-lengths (mean temperature, 59·8° C.).

Wave-length in $\mu\mu$ . . .	185	186	198	206	214	219	257	274	298	346	361	441	480	503
$\delta\mu\times10^5$ . . .	+2·319	+2·271	+1·965	+1·832	+1·728	+1·666	+1·374	+1·301	+1·225	+1·141	+1·127	+1·041	+1·020	+1·021

References.—A great deal of information is to be found in various journals on optical glass. Much of the work that has been done on Jena glasses in particular has been collected by Hovestadt, whose book has been translated into English by Professor and Miss A. Everett, and published under the title *Jena Glass*, and reference to this book is recommended for further information. For recent work, reference should be made to the *Journal of the Society of Glass Technology*, both for special articles and for information on material published elsewhere.

Table II. gives the results of a series of measurements on typical specimens of optical glass manufactured by Schott & Gen., Jena (H. Th. Simon, *Ann. d. Phys. u. Chem.*, 1894, liii. 555).

The results of measurements on the temperature coefficients of refractive index in the case of transparent crystals vary considerably. The values obtained by a number of observers at various mean temperatures are given in Table III. (Landolt-Bornstein Tables, 4th ed., 1912, p. 976).

The "flow temperature" given in Table IV. is the minimum temperature at which the glass is in a fluid state. It depends on the duration of the heating process and on the size of the specimen. The following values are for specimens in the form of cubes, the length of the side of which is 25 mm. The results refer to specimens of optical glass manufactured by the Sendlinger Optische Werke, Zehlendorf, near Berlin. In the first column the figures refer to the refractive indices and ν values of the glasses. Thus 510/634 shows that the specimen has a refractive index (for sodium D) of 1·510 and a ν value of 63·4 (F. Weidert and G. Berndt, *Z. f. tech. Phys.*, 1920, No. 6).

TABLE IV

MELTING TEMPERATURES OF SPECIMENS OF OPTICAL GLASS

Type of Glass.	Analogous Schott Type.	Flow Temperature in ° C. for three times of Heating.		
		½ Hour.	2 Hours.	6 Hours.
Kron 510/634	O. 144	850	815	775
Kron 516/640	O. 3832	810	795	780
Flint 549/461	O. 378	740	725	685
Barion 573/575	O. 211	910	885	860
Barint 580/538	O. 722	845	805	785
Barion 590/612	O. 2122	845	830	795
Barion 609/589	O. 2071	870	835	820
Flint 613/369	O. 118	730	695	680
Barion 614/564	O. 2994	840	815	800
Barint 626/393	O. 748	780	730	685
Flint 649/338	O. 102	660	645	630

Table V. gives the tensile and compressive strengths of different types of optical glass (Landolt-Börnstein Tables, 4th ed., 1912, p. 54). For the results of more recent measurements reference should be made to G. Berndt, *Verh. d. D. Phys. Ges.*, 1917, xix. 314; *Z. f. Instrumentenkunde*, 1920, xl. 20, 37, 56, 70. For the elastic constants of transparent crystals see Landolt-Börnstein Tables, 4th ed., 1912, p. 50.

TABLE V

ELASTIC CONSTANTS OF SPECIMENS OF OPTICAL GLASS

Type of Glass.	Tensile Strength. kgs./mm².	Compressive Strength. kgs./mm².	Ratio of Compressive Strength to Tensile Strength.
Soda alumina borosilicate	6·76	126·4	18·7
Densest lead silicate	3·28	60·6	18·5
Alumina lead borosilicate	5·66	105·7	18·7
Soda alumina borate	4·93	81·2	16·5
Baryta zinc borosilicate	7·21	84·0	11·7
Dense potash lead silicate	6·01	77·5	12·9
Soda zinc silicate	7·84	97·8	12·5
Potash alumina phosphate	5·46	71·7	13·1
Potash baryta soda silicate	6·09	91·6	15·0
Soda lead zinc silicate	6·42	99·0	15·4
Potash lime silicate	7·52	68·3	9·1
Baryta alumina phosphate	7·42	75·0	10·1
Potash zinc silicate	8·09	73·9	9·1
Dense lead silicate	4·97	67·3	13·5
Soda lime zinc silicate	7·46	112·9	15·1

T. S.
J. S. A.

OPTICAL GLASS, MANUFACTURE OF. See "Glass," § (17).

OPTICAL IMAGE: ITS DEFECTS. See "Optical Parts, The Working of," § (3) (i.).

OPTICAL PARTS, THE WORKING OF

§ (1) INTRODUCTION.—For the production of an optical instrument the collaboration of the designer, the computor, and the optical worker is essential. Particulars of the function of the proposed instrument and the conditions to be fulfilled having been determined, the designer is able to prepare a general scheme of the mechanical and optical arrangement. So far as the optical portion is concerned, this involves a general knowledge of the limitations of computation and the accuracy attainable in the practical working of the parts.

Where possible, the designer avoids the introduction of extreme angular apertures that might necessitate the use of very special types of glass or increase the computational difficulties. But the tendency must always be towards the imposition of increasingly drastic demands, and frequently for one or more details of the optical system the designer must finally rely upon the utmost skill of the computor. Close collaboration between the computor and the workshop is also essential. Thus, for example, it may be desirable to use particular types of glass that happen to be in stock or standard test plates and tools.

It is the function of the optical worker, with whom this article is mostly concerned, to form the parts to the specified dimensions, to polish the surfaces, to examine the performance of the finished work, and when possible, to compensate the defects whether of surface or substance.

§ (2) METHODS.—As to the methods employed in the workshop, much depends upon the character of the work. Large astronomical objectives for which the demand is extremely small are invariably produced singly and involve the exercise of very special craftsmanship.

One renowned manufacturer[1] has stated as regards design that object-glasses cannot be made on paper. The empirical method of producing such parts consists in the computation of the objective by means of the simplest formulae with a view principally to the elimination of chromatism, and the determination of the desired focal length, and then in the removal of residual aberrations, and more particularly spherical aberration, by mechanical local retouching based upon optical examinations of the images formed by different zones and portions of the objective. From the point of view of the perfection of the objective this empirical method appears to be justified by the excellence of the results attained. Other makers of large astronomical object-glasses[2] advocate the alternative method. By rigid computation they determine the curves, thicknesses, separations, and apertures of the parts, and in the workshop they endeavour to attain the desired degrees of freedom from the several aberrations by the least possible departures from the calculated data. Local retouching, however, can hardly be avoided in the production of large optical parts. Peripheral or central zones of one or more of the surfaces may have to be retouched in order to produce an aspherical surface for the compensation of spherical aberration. Regular cylindrical retouching may be necessary for the correction of astigmatism. Surface imperfections, defects of homogeneity, defective annealing, constraint during working processes, and thermal effects may all involve irregular retouching.

But the organisation of most optical workshops is arranged principally for the production of moderately small lenses and prisms, of which comparatively large quantities have usually to be produced, as, for example, in the manufacture of binoculars and cameras. For reasons of cost, retouching of such parts is not permissible, and indeed rarely necessary, if the materials are well selected. They must be produced within limits specified by the

[1] H. Grubb, "The Production and Testing of Telescope Objectives and Mirrors," *Nature*, 1886, xxxiv. 85.

[2] S. Czapski, *Zeits. f. Instk.*, 1887, vii. 101.

designer and computor, and defective elements are rejected. Trial and error methods play but little part in such an organisation. Each step is based upon precise measurements.

There is an intermediate class of multiple work involving parts the size and cost of which are such as to make a certain amount of retouching necessary, although such retouching must always be regarded as undesirable.

At the present time the technique of the optician is in advance of that of the glass-maker. By machinery it is possible to produce surfaces that may be regarded as optically perfect. A prism, such as is used at the ends of a rangefinder, having perfect optical surfaces will not necessarily give a well-defined image, owing to small imperfections of the glass, such as defective annealing or homogeneity. The larger the piece the greater is the difficulty in obtaining glass that is sufficiently homogeneous to suit the requirements of the optician.

In such work it is customary to form and polish the pieces by machine, to test the perfection of the definition of the parts, and, if necessary, of their individual surfaces. Those that fail to pass and that exhibit certain types of imperfect definition may be saved by hand retouching, or, in the more regular cases, by machine. Of the remaining defective pieces, some may require re-annealing. The others must be finally rejected unless they can be utilised in instruments of lower power or inferior quality.

There is another class of work, such, for example, as condenser lenses and to some extent spectacle lenses, in which no great accuracy either of form or surface is required, and which will not be further considered.

It will be evident from the above general remarks that the production of high-quality optical parts cannot be reduced entirely to purely mechanical operations. A certain amount of hand work is involved, not only because of the need for the irregular touching of surfaces, but also for the production of certain delicate and accurate details for which suitable machinery has not yet been elaborated.

In the most highly organised workshops there may therefore still be seen in operation the early hand processes elaborated by such pioneer workers as Huyghens, Hooke, Newton, Father Cherubin, Herschel, and Molyneux.

§ (3) THE OPTICAL IMAGE. (i.) *Its Defects.*—Whether for the purpose of final or work-in-progress inspection, or for retouching, it is necessary to analyse the definition of the image produced by the optical part or system in question, and to diagnose the probable defects with a view to their possible correction.

It is not sufficient to make the broad statement that the definition is bad, because it may be possible to compensate the defect in the case of large and costly parts without much additional expense. Suppose the optical part to be examined is a thick parallel plate of glass whose transmission surfaces have been optically worked and may be regarded as being unquestionably true, and suppose the part is placed in the parallel beam before the objective of a telescope which is directed upon a collimator or test image comprising small holes or lines. If the glass is perfect, its insertion in the path of the light should not affect the appearance of the collimator image. If, to obtain a sharp image, a readjustment of the telescope eyepiece is necessary, the parallel plate has a focus error. It is equivalent in its action to a very weak lens. Examination by means of the interferometer may reveal a regular change of optical density from the centre outwards as represented by the circular bands of *Fig.* 1 (*a*). The defect in question may be due to imperfect annealing, or a regular defect of homogeneity. Double refraction resulting from imperfect annealing of the glass may be tested by inserting a half-wave plate and observing the extent to which the bands are displaced as the half-wave plate is rotated.

If the defect is due to internal stresses, the glass should be re-annealed, but if the substance itself is at fault the definition may be corrected by making one of the surfaces spherical. In the case of a prism a transmission face would be chosen for the operation in preference to a reflecting surface, since oblique reflection at a curved surface necessarily introduces astigmatism, which again might require to be compensated by working one of the transmission faces to a suitably oriented cylindrical form.

If the image appears sharply defined but oval shaped, the defect is one of astigmatism. The interferometer appearance of the glass may be as indicated in *Fig.* 1 (*b*). A cylindrical surface suitably oriented will suffice to compensate the defect, but if it is the glass that is strained it should be re-annealed.

Instead of a single clearly defined image a complex multiple image may be produced. This defect is due to heterogeneity of the glass, the structure of which when viewed by means of the interferometer may appear as in *Fig.* 1 (*c*). Such a structure might result from imperfect solution of the constituents, the nucleus of a sphere, in a soda lime glass, for instance, being quartz, having an ordinary refractive index of 1·54, and the various layers being silicates of gradually diminishing refractive index, approximating to about 1·5, which is the index for sodium silicate. Each sphere acts as a small aperture lens and forms a separate part of the multiple image. Local retouching may sometimes suffice for

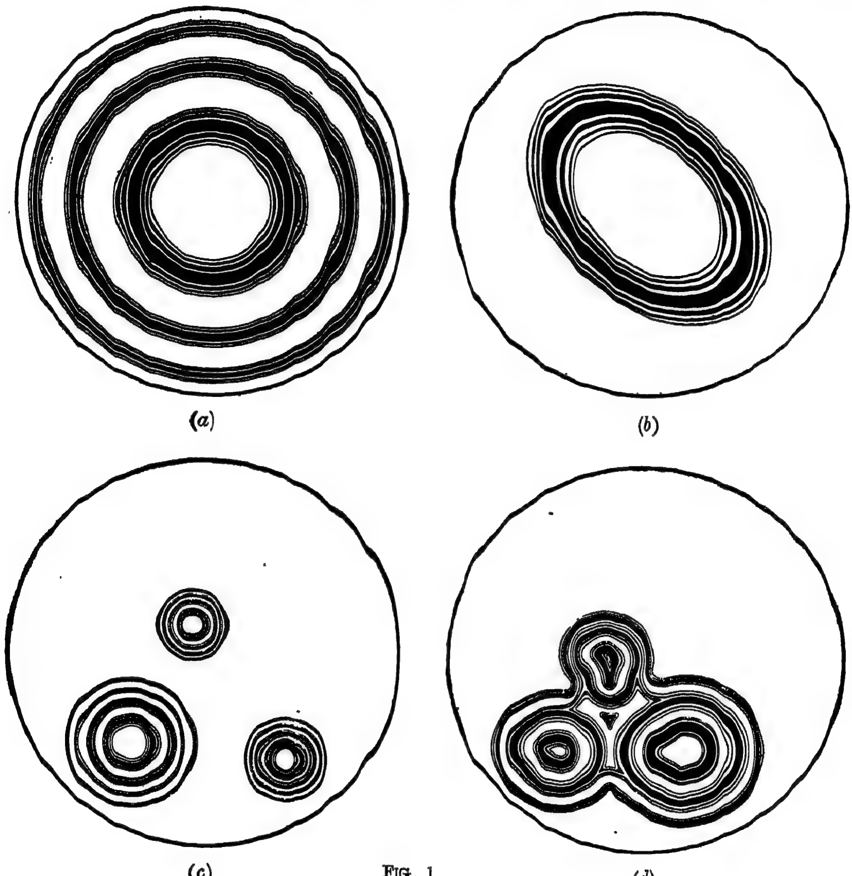
the compensation of these multiple image defects.

If the spheres are distorted and merge into one another as indicated by the interferometer appearance in *Fig.* 1 (*d*), the image produced will be indefinite or "fuzzy." Retouching in such a case is usually impracticable and the part must accordingly be finally rejected.

These several defects may appear singly or in combination and in various degrees. Thus, for example, in practice pure astigmatism is seldom found in optically parallel plates of thick glass. Usually the appearance of the image is attributable to lenticular and astigmatic defects which may be compensated by working one surface to a suitable cylindrical and the other to a spherical form.

It should be remembered that in the estimation of the various types of definition it is necessary that the inspector should know to what extent the errors are attributable to the defects of his eye, or to the apparatus employed, which should involve the smallest possible number of optical parts.

Veins, if not numerous, and if unaccompanied by other defects, have no serious effect upon the definition. A vein is usually an extremely fine thread of glass richer in silicate of alumina than the surrounding substance. Each vein is equivalent to a long astigmatic lens of extremely short focus, and the separate image formed by it is usually so diffused as to be invisible. But if the test object is a fine line, then by a movement of the eye the image may be momentarily occulted by the passage of the vein in front of it. A number of parallel veins may give rise to a rippling appearance of the image, and be sufficient cause for the rejection of the part. Often, however, the veins in glass are an indication of defective homogeneity, and in the best work the use of such glass should be avoided.



(*a*) (*b*)

(*c*) FIG. 1. (*d*)

(ii.) *Spherical Aberration.*—For the investigation of spherical aberration errors

Foucault's[1] original knife-edge method is still frequently employed, especially in the case of large object-glasses. If the eye of the observer is placed close to the focal point of the object-glass and receives all the rays falling upon it from a star or its equivalent, the whole aperture of the object-glass will appear illuminated. If then the focus is cut transversely by means of a knife edge, which should be mounted upon a fine operating screw, the whole aperture will become dark at the instant the knife edge passes the axis, if the objective is ideally perfect. If all the rays from the objective do not pass through the ideal focus, a number of rays will escape past the edge and still reach the eye, and the aperture will be partially illuminated, the distribution of the light affording an indication of the extent and distribution of the aberration. In the direct focussing method[2] the eye receives the focal image, formed by the objective, of a natural or artificial star, which appears as a system of diffraction rings surrounding a central spot of light. By examining the image out of focus on the near and far sides, the nature of the spherical aberration either over the whole surface or over zones may be deduced from the change in the distribution and character of the rings. Irregular defects of the surfaces or of the glass may be indicated by distortion of the rings. Thus an oval-shaped system would denote astigmatism. Chromatic aberrations may also be investigated by using homogeneous light of various wave-lengths. These two methods, and particularly the former, give qualitative results rather than quantitative. The latter method in the hands of a skilled observer is more precise and the appearances may be more directly interpreted.

Special forms of Michelson's Interferometer[3] are now frequently employed for the examination of prisms and lenses, and even of complex optical systems, although the interpretation of the results is then difficult. When a simple prism or object-glass is examined in this way the surface will appear uniformly illuminated if the piece under test causes no relative retardation of any portion of a plane wave front, assuming, of course, that the optical parts of the interferometer itself have been accurately compensated.

A concentric ring appearance would indicate curvature of one or more of the surfaces or a regular variation of the material. Irregularities of surfaces or substance would be indicated by a distortion of the system of rings.

The interferometer does not discriminate between retardations attributable to the glass and to the surfaces, except those due to defective annealing, which may be detected by observing the shift of the rings as a half-wave plate inserted before the eye is rotated. It is the resultant retardation that is indicated by the interferometer, and by the use of the interferometer alone it is not possible to allocate the component defects to the particular elements. But in practice it is sufficient to compensate the resultant defect by retouching one of the surfaces, and for this purpose a contour of the ring system as seen by the observer is painted upon the surface selected for the retouching operation. By polishing away the high portions and repeating the observations a comparatively uniform distribution may ultimately be obtained, and the definition be thereby greatly improved. Only large prisms and objectives can be subjected to an expensive process of this kind. For the production of aspherical surfaces the method is particularly valuable.

(iii.) *Chromatic Aberration.* — Chromatic aberrations may be investigated by the Vogel[4] method, provided the aperture is not so small that a large shift of the eyepiece is necessary to detect a difference of focus. It is also necessary that the chromatic effects due to the eye and the eyepiece or other parts involved should be previously determined.

In the eyepiece there is mounted a direct vision prism which forms a spectrum of the star image formed in the field of view through the intermediary of the part to be tested. If the correction is perfect, the spectrum, when the eyepiece is focussed, will appear as a thin line of uniform width, but if there is any chromatic aberration of particular colours the corresponding portions of the spectrum will be broader than the remaining parts. A measurement of the aberrational defects may then be obtained by observing the shift of the eyepiece necessary to reduce the various widths to a minimum amount. This method indicates only the general chromatic defect. It can hardly discriminate between zones, and if the aperture is reduced by stopping out the peripheral portions, the usefulness of the method is diminished.

Precise measurements of the various aberrations are obtainable even in the case of objectives of small aperture by the Abbe Focimeter method.[5]

But probably the most complete and precise investigation of the spherical and chromatic aberrations of any optical part is obtainable

[1] Foucault's method is discussed in the handbook *On the Adjustment and Testing of Telescope Objectives*, by T. Cooke & Sons, 1896.

[2] *On the Adjustment and Testing of Telescope Objectives*, T. Cooke & Sons, 1896.

[3] F. Twyman, *On the Use of the Interferometer for Testing Optical Systems*, Traill Taylor Lecture, Roy. Phot. Soc., 1918; *Phil. Mag.* vol. xxxv., Jan. 1918; "Correction of Optical Surfaces," *Astro-Phys. Journal*, xlvii., No. 4, 1918; Michelson, *Astro-Phys. Journal*, June 1918.

[4] B. C. Vogel, *Monatsberichte d. Berl. Ak.*, 1880.

[5] S. Czapski, "Methode und Apparate zur Bestimmung von Brennweiten nach Abbe," *Zeitsch. f. Instrk.*, 1892, xii. 185.

by the Hartmann method [1] by means of which the final positions of particular rays are located in a manner comparable with the trigonometrical computation of the paths of rays. Symmetrically arranged holes in a diaphragm determine the portions of the optical part to be traversed by the fine bundles of rays which are finally received upon photographic plates situated at symmetrical positions before and after the focal plane. From the spacing of the images on the photographic plates in comparison with the spacing of the holes in the diaphragm the aberrations can be determined. Separate observations with light of different colours are made for the determination of chromatic aberration. The original method of examining the optical performance of an objective by means of selected rays is described by Father Cherubin,[2] who also demonstrates the effects of defective centring and emphasises the importance of accuracy in this respect.

From a consideration of the light intensity at a focal point Lord Rayleigh [3] has stated that "in general we may say that aberration is unimportant when it nowhere (or, at any rate, over a relatively small area only) exceeds a small fraction of the wave-length (λ). Thus in estimating the intensity at a focal point where in the absence of aberrations all the secondary waves would have exactly the same phase, we see that an aberration nowhere exceeding $\frac{1}{4}\lambda$ can have but little effect," and again, "An important practical question is the amount of error admissible in optical surfaces. In the case of a mirror reflecting at nearly perpendicular incidence there should be no deviation from truth (over an appreciable area) of more than $\frac{1}{8}\lambda$. For glass, $\mu - 1 = \frac{1}{2}$ nearly; and hence the admissible error in a refracting surface of that material is four times as great." This estimated value of Lord Rayleigh of the permissible phase difference is generally confirmed by practical experience.

§ (4) Test Plates.—Test plates are very commonly used for the examination of the surfaces not only during the progress of the work, but also in the final testing, with a view to the allocation of any observed defects of definition.

Sir Isaac Newton [4] has dealt very exhaustively with the colours of thin plates, which had previously been observed, but not completely described by other workers. Although it is evident from Newton's description that he realised that these colours afforded an indication of minute irregularities of thickness of the intervening layer, it is not so clear that it was his practice to test the surface being operated upon in comparison with a test plate having a very perfect optically flat surface.

These test plates which are the most valuable of the practical optician's appliances are usually made of quartz. But for reasons of economy glass test plates which from time to time must be compared with the standard quartz plates are employed. Quartz has the primary advantage of hardness. The frequent cleaning of the test surface that is essential soon impairs the definition in the case of a glass plate, and even in the case of quartz a good worker handles the surfaces with the greatest care.

For flat work the parallel quartz plate, having an approximate thickness of about one-fifth its diameter, is polished on both sides and worked as perfectly flat as possible on at least one side. When the test plate surface is placed upon the surface to be tested, if the latter is the larger and the more rigid, a thin layer of air is enclosed between them. Diffused white light which is reflected from the two surfaces of the air film gives rise to interference colours which may be viewed by a suitably placed eye. If monochromatic light is used numerous fine interference lines are visible, indicating minute variations that cannot be detected when using white light. Since the temperatures of the work and of the test plate will most probably be unequal, several systems of Newton's rings more or less distinct will be visible. As the temperatures equalise the rings will broaden, and in the course of 5 to 30 minutes, or more according to the volume of the parts, if the surface is perfect, a uniform straw yellow appearance may be obtained by skilful manipulation of the plate. If the surface is regularly curved concentric rings will be seen, and the greater the curvature the closer will be the spacing of the rings, provided the surfaces are clean and dry and the test plate is properly handled. Distorted rings or bands indicate irregularity of the surface relatively to the test plate. If when the test plate is gently pressed eccentrically by means of a pointed piece of wood, and not the finger, the centre of the rings moves towards the point of application; the surface under test is convex. It is very important that the adjacent surfaces of the work and the test plate should not only be thoroughly clean and free from the minutest specks of dust, but that they should also be free from any trace of moisture, the capillary action of which would locally distort the pieces and give a false indication.

[1] J. Hartmann, "Objektivuntersuchungen," *Zeitsch. f. Instrk.*, 1904, xxiv.; H. Fassbender, *Zeitsch. f. Instrk.*, 1913, xxxiii. 177.

[2] Père Cherubin, *Vision parfaite*, ii. 109, also p. 25; William Molyneux, *Dioptrics*, 1692, part ii. chap. iv. p. 222.

[3] Lord Rayleigh, *Scientific Papers*, iii. 100, 104; also L. Silberstein, "Light Distribution round the Focus of a Lens," *Phil. Mag.*, 1918, xxiv. 35.

[4] Sir Isaac Newton, *Opticks*, 1704, book ii. part i.

For the best quality of work great skill is necessary in the use of a test plate, the indications of which require careful interpretation. Flat surfaces may be finally tested by an analysis of the reflected image in the manner described in § (2). The reflection test is the more reliable, but the test plate has the merit of great convenience, and for the greater part of the work the optician is called upon to perform it is thoroughly reliable and invaluable. For curved work one surface of the test plate is worked to the appropriate curve and the other surface is made flat. The radius of the curved test surface is measured by means of a spherometer or the radius or focus may be determined optically. It will be understood that the primary function of the test plate is not to obtain absolute measurements, but to determine the extent and nature of the difference between the worked surface and the test plate. To indicate the permissible amount of irregular distortion is hardly possible. If there is any noticeable irregularity the surface should be reworked. Slight regular ellipticity may be permissible when the astigmatisms of two transmission faces are normal to one another. Curvature to the extent of one and a half to two complete rings is generally regarded as being just permissible. It might be thought that such a statement would require extensive qualification according to the size of the part, its function, and its position in the optical system, particularly as regards its distance from a focal plane. Practical experience, however, shows that this limit of two rings, which accords with the practice of at least one large German firm,[1] covers a very wide range of work. Factory conditions render it practically impossible to adjust the limits to suit the requirements of individual parts. The tendency is towards the adoption of approximately one general and high standard of optical quality.

§ (5) Optical Processes.—For practical reasons it is customary to divide the process of working optical surfaces into three stages: (1) The forming stage in which the size and shape of the part is accurately determined; (2) the smoothing stage; and (3) the polishing stage.

Stages (1) and (2) involve the use of abrasives and these operations are accordingly performed in rooms quite separate from the polishing departments, where fine media only are employed. But it must not be assumed that the polished appearance of the surface as distinct from regularity of surface is a phenomenon that only takes place as a result of the polishing process. Actually a certain amount of polish is associated with the use of the roughest abrasive.

[1] W. Zschokke, *Festschrift*, Firma C. P. Goerz, p. 140.

It is only within comparatively recent years that the molecular regularity of polished surfaces has received recognition. The earlier conception is very clearly expressed by Sir Isaac Newton [2] in the following words:

"For in polishing glass with sand, putty, or tripoly, it is not to be imagined that these substances can, by grating and fretting the glass, bring all its least particles to an accurate polish, so that all their surfaces shall be truly plain or truly spherical and look all the same way so as together to compose one even surface. The smaller the particles of those substances are the smaller will be the scratches by which they continually fret and wear away the glass until it be polished, but be they never so small they can wear away the glass no otherwise than by grating and scratching it and breaking the protuberances, and, therefore, polish it no otherwise than by bringing its roughness to a very fine grain so that the scratches and frettings of the surface become too small to be visible."

There is no indication here that Newton regarded the surfaces as being molecularly regular or comparable with the surface of a liquid.

Subsequent writers have not hesitated to accept without question and to repeat the statement of so authoritative an observer as Newton. Thus, for example, Coddington,[3] Sir J. F. W. Herschel,[4] and Sir David Brewster [5] use practically the words of Newton when describing a polished surface.

If polishing were merely a continuation of the grating and fretting of the surface protuberances it should be possible to observe a continuous sequence of appearances from coarse conchoidal fractures to a grain of ultramicroscopic character. But if the operation of polishing a smoothed surface is performed under the microscope it will be seen that numerous patches of perfect polish akin to the still surface of a liquid are formed almost instantly, and that these patches exhibit no intermediate structure other than accidental scratches and kindred defects.

Lord Rayleigh [6] in a lecture on "Polish" appears to have been the first to describe their appearance. He states that, "In view of these phenomena we recognise it is something of an accident that polishing processes as distinct from grinding are needed at all, and we may be tempted to infer that there is no

[2] Newton's *Opticks*, 1704, second book, p. 68.
[3] Coddington's *Optics*, 1825, p. 32.
[4] Sir J. F. W. Herschel, *Encyclopaedia Metropolitana*, 1830, "Optics," p. 447.
[5] Sir David Brewster, *Optics*, p. 159.
[6] Lord Rayleigh, "Polish," Royal Institution, March 29, 1901. See also Lord Rayleigh, "Interference Bands and their Applications," Royal Institution, March 24, 1903. Also "Polishing of Glass Surfaces," *Proc. Opt. Convention*, No. 1, 1905, p. 73.

essential difference between the operations. This appears to have been the opinion of Herschel (as expressed in the *Enc. Met.*, art. 'Light,'[1] pp. 447 to 830), whom we may regard as one of the first authorities on such a subject. But although perhaps no sure conclusion can be demonstrated, the balance of evidence appears to point in the opposite direction." . . . "Under those conditions which preclude more than a moderate pressure it seems probable that no grits are formed by the breaking out of fragments but that the material is worn away almost molecularly." . . . And later he states: "But so much discontinuity as compared with the grinding action has to be admitted in any case that one is inevitably led to the conclusion that in all probability the operation is a molecular one and that no coherent fragments containing a large number of molecules are broken out. If this were so there would be much less difference than Herschel thought between the surfaces of a polished solid and of a liquid."

Although the molecular character of the polishing operation and the similarity of the surface produced to that of a liquid are quite definitely expressed, and although Lord Rayleigh has referred in other of his papers to the remarkable pool-like appearance of elementary polished patches of a glass surface, it is not quite clear whether he regarded the result as being due to the removal of the substance molecule by molecule as distinct from the removal of minute aggregates of molecules or as being due to a molecular rearrangement or flow of the surface molecules as in the case of a liquid.

This latter conception is attributable very largely, if not entirely, to Sir George Beilby, who has developed it in a series of papers[2] dealing principally with metal surfaces, the tenacity of which is such that the surface amorphous layer is capable of bridging over surface cavities even when these are not completely filled with debris. It is very doubtful if any such bridging over of even minute cavities occurs in glass owing to the small cohesion of the silicates as compared with that of the metals.[3]

According to the molecular flow[4] theory of polishing, the forces exercised by the polisher upon the surface molecules of the glass suffice to overcome the cohesive forces binding them together, with the result that the molecules rearrange themselves uniformly under the action of their surface tension forces. Thus it would appear that the grain of a polished surface, being molecular, is much finer than is actually required for the regular reflection of even the shortest visual rays at normal incidence.

Polished layers may be produced in several ways although in all cases the action is fundamentally the same. Fire-glazed surfaces result from the thermal agitation and consequent flow of the surface molecules. Chemical forces produce a similar result. Provided precautions are taken to prevent the accumulation of fluoride crystals, very perfect light-reflecting glass surfaces may be produced by the action of hydrofluoric acid.[5]

When a piece of glass is fractured the forces at the cleavage edge so profoundly disturb the molecules that they are able to flow and form the characteristic polished appearance of a fractured surface.

It is hardly possible to fracture a piece of glass so suddenly that its surface is not polished. Under the microscope a rough ground surface is seen to consist of numerous conchoidal depressions the surfaces of which are all light reflecting, and which may indeed be made to act as so many separate lenses of poor quality.[6] Closer examination will disclose a rounding of all the ridges where the conchoidal surfaces intersect that can only be attributable to viscous flow. The ridges have a characteristic yellow-green colour whether the glass is flint or crown. Where these ridges intersect one another elementary polished patches are found, and if an attempt were made to polish a rough ground surface of this type, it would be seen that these elementary patches would be extended until they joined one another with the ultimate formation of a continuous polished surface.

§ (6) Abrasion.—In order that the workshop processes may be more fully understood, the features characteristic of abrasion and polishing must be considered in detail.

Suppose a small steel ball is pressed upon the polished surface of a glass cube the transverse faces of which are also polished so that the stresses introduced may be viewed by means of a polariscope.[7] When the pressure

[1] Herschel's description is practically a repetition of Newton's earlier observations.

[2] *Papers*, G. T. Beilby: "Surface Flow in Crystalline Solids under Mechanical Disturbances," *Proc. Roy. Soc.*, 1903, lxxii. 72; "The Effects of Heat and of Solvents on Thin Films of Metal," *Proc. Roy. Soc.*, 1903, lxxii.; "The Hard and Soft States in Metals," *Phil. Mag.*, Aug. 1904; "The Influence of Phase Changes on Tenacity of Ductile Metals," etc., *Proc. Roy. Soc.*, 1905, A, lxxvi.; "The Hard and Soft States in Ductile Metals," 1907, A, lxxix.; "Surface Flow in Calcite," *Proc. Roy. Soc.*, 1907, A, lxxxii.; "Transparence or Translucence of the Surface Film produced in Polished Metals," *Proc. Roy. Soc.*, 1914, A, lxxxix.

[3] "Some Notes on Glass Grinding and Polishing," J. W. French, *Opt. Soc.*, 1916, xvii. No. 2.

[4] See also "Solids, The Flow of," Vol. V.

[5] Lord Rayleigh, "Polish," Royal Inst., March 29, 1901. The writer has confirmed Lord Rayleigh's experiments, and has reduced by means of hydrofluoric acid the surface of a polished plate to the extent of about 50 μ without adversely affecting its optical perfection.

[6] Lord Rayleigh, "Interference Bands and their Applications," Royal Inst., March 24, 1893.

[7] J. W. French, "Percussion Figures in Isotropic Solids," *Nature*, Nov. 20, 1919, p. 312.

is very light the appearance between the crossed Nicols is as in *Fig.* 2. The central black cone has an angle of about 20°, which appears to be practically independent of the pressure. The cone of strain *b*, *b* has an angle of about 90°. Some surface light is visible at *d*, *d*. At low pressures the dark cones *c* and *a* merge softly into *b*. As the pressure is increased the interfaces become more intensely defined, but the angles do not alter appreciably. Further gentle increase of pressure causes the surface layer to rupture as in *Fig.* 3, which is a photo-micrograph of an etched polished surface repeatedly ruptured by gentle impact. If the Nicols are paralleled black rays will be seen proceeding from the edge of the crack as in *Fig.* 4, their direction indicating that the crack is normal to the surface. The character of *Fig.* 2 is not appreciably altered.

A new phenomenon makes its appearance when the pressure is again increased. Immediately under the ball there appears as in *Fig.* 5 a sphere pierced by the filament of the cone *a*, and having a black outline tinged with red on the outside. The interior is filled with green-blue light; otherwise the general appearance of *Fig.* 2 is unaltered. If now the Nicols are paralleled, it will be seen that the well-known type of conical fracture [1] has been produced as indicated in *Fig.* 6. The coloured sphere of *Fig.* 5 is a certain indication of the existence of a cone fracture. Examination with the Nicols in an intermediate position shows that the cone fracture which follows the surface of *b* is tangential to the sphere which covers it. If the pressure is again increased the crushing point is soon reached. The glass under the ball collapses almost explosively, a distinct click being audible, and the ball sinks deeply through the surface. At the moment of this collapse, the cone of light *b* broadens out owing to the extension of the area of pressure. The cone fracture also may extend horizontally like the brim of a hat, thus definitely terminating the depth below the surface, and the space within the cone becomes cleft by one or by two fracture planes normal to one another and having their line of intersection on the axis of the dark cone *a*, as in *Fig.* 7. The diametral planes may be extended to the brim as in *Fig.* 8. Under crossed Nicols two new small coloured spheres indicative of subsidiary planes may be observed as in *Fig.* 9.

Now the grains of an abrasive such as carborundum are nodular in parts and hard and

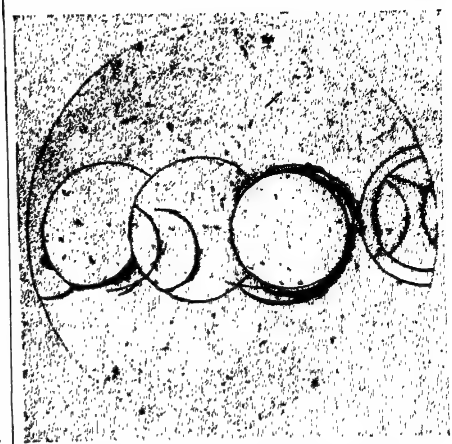


FIG. 3.—Surface etched after rupturing. Magnification 40 diameters.

akin to the steel balls used in the experiment. But a rough ground or smoothed glass plate exhibits practically no cone fractures, notwithstanding the vast number of impacts that must have occurred in the operation. Only if the tool carrying the abrasive is lowered sharply on to the glass surface will a number of cone fractures probably be formed.

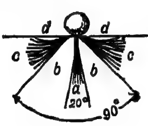


FIG. 2.

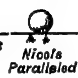


FIG 4.

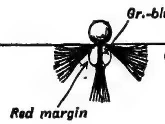


FIG. 5.

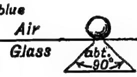


FIG. 6.

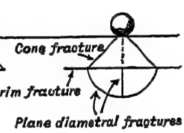


FIG. 7.

From this it seems evident that the grinding of glass is not the result of any such normal pressure of the grains, and another explanation must be obtained. In the experiments with the steel ball as described above the forces were symmetrical about the vertical axis. When the pressure is applied near the edge of the surface, the new appearance corresponding with the stage illustrated in *Fig.* 2 will be as illustrated in *Fig.* 10, from which it will be

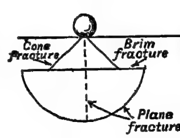


FIG. 8.

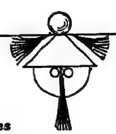


FIG. 9.

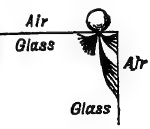


FIG. 10.

[1] "L'Éclatement," Ch. de Fréminville, *Revue de Métallurgie*, Sept. 1914, sect. vii. p. 66.

seen that the central cone is now deviated towards the side. Its axis along which fracture finally takes place follows the characteristic conchoidal section. It is presumably the impact of the abrasive grains on the edges of cavities that produces the conchoidal splintering of a ground surface as indicated in *Fig.* 11.

As it is the transverse movement of the tool relatively to the glass that forces the hard grains against the edges of the cavities, it is to be expected that the rate of grinding will depend upon the speed of movement of the tool relatively to the work, and also upon the pressure exerted by the grains. This is confirmed, so far as the effect of relative speed is concerned, by the results of carborundum abrasive tests which indicate that the weight of glass removed is directly proportional to the relative speed,[1] the other conditions being maintained constant. To eliminate the effect of loss of cut fresh abrasive must be applied at intervals of about one minute, and precautions must be taken to ensure constancy of the general conditions throughout the tests.

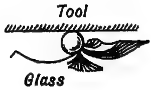


FIG. 11.

From an examination of the previous *Fig.* 11, it will be seen that the larger the grain the deeper, in general, will be the point of impact and the larger the splinters removed. The rate of grinding is therefore dependent also upon the size of the grain. Thus by reducing the coarseness of the abrasive at each stage of the grinding operation, a surface of any desired degree of fineness may be obtained; and indeed the surface may be highly polished by means of the same material as is used for the rough grinding, provided a sufficiently fine grain of a suitable character is obtainable. Prior to the time of Antheaulme[2] it was the practice of the earlier opticians[3] to make the polishing process merely a continuation of the grinding stage, the sand first used being ground down sufficiently in the operation to serve as a polishing medium in the later stages.

Highly polished optical surfaces have also been produced by the use of very fine grades of carborundum. As much more suitable polishing media than fine carborundum are available, this experiment is mainly of interest as emphasising that there is no strict line of demarcation between abrasion and polishing. The removal of material and the production of the amorphous polished layer occur simultaneously, although to different extents. Thus in the coarsest grinding there is removal of material and only very slight surface flow along the ridges between the concavities of the surface, since the cleavage flow over the depressed surfaces of the conchoidal fractures does not contribute except in the last stage to the final surface.[4] As the abrasive becomes finer the splinters decrease in size and the material removed diminishes, while the amount of surface flow over the network of ridges increases.

If in the grinding process the flat grinding tool is so hard that it rides over the grains, the impacts would be more normal to the surface and undesirable cone-fracturing would occur. The finer the abrasive the harder may be the smoothing tool. Lead, zinc, copper, aluminium, brass, and cast iron have all been used for various kinds of work and various abrasives, but fine-grained cast iron and brass, free from surface defects, are most generally, and indeed almost universally, employed. Steel is too hard for even the finest grades of abrasive.

After the surface has been ground to the necessary degree of fineness, polishing may be commenced, the purpose of the process being to promote the greatest possible amount of surface flow while avoiding the conchoidal splintering characteristic of the grinding process.

A very small amount of material is removed during the polishing action, but the nature of the abrasion, if it can be so termed, is characteristically different from that during grinding. Minute grooves are ploughed through the amorphous surface layer, and small portions of the amorphous substance become disengaged by the action of the polisher and are removed. These fragments may be recovered by dissolving away the rouge and resinous substances. The residue has a sparkling snow-like appearance and consists of extremely minute unresolvable particles cemented loosely together possibly by surface fusion along their edges.

§ (7) POLISHING.—A clear distinction must be drawn between a polished surface and one that is at the same time optically regular. This will be more easily understood by considering the action of a cloth polisher as compared with that of a pitch polisher.

Suppose in *Fig.* 12, A is the fine ground surface greatly magnified composed of flat elementary areas with numerous conchoidal depressions. A surface of this kind is said to be grey, the appearance being due to irregular reflection or scattering of the light. B is a layer of soft cloth or felt cemented to the regular surface of the metal runner C, which may be of aluminium, brass, or iron. Over the flat areas of A the drag, and consequently the reduction of the general level, will be greatest. Over the surfaces of the

[1] "More Notes on Glass Grinding and Polishing," J. W. French, *Trans. Opt. Soc.*, Jan. 1917, xviii.

[2] *Histoire des mathématiques*, Montucla, vol. iii. part v. book ii. p. 498.

[3] Renati Descartes *Opera Philosophica*, "Dioptricis," 1656, cap. x.; Johannes Zahn, *Oculus Artificialis Telediopticus*, 1702.

[4] "Some Notes on Glass Grinding and Polishing," J. W. French, *Trans. Opt. Soc.*, Nov. 1916, xvii. No. 2.

depressed areas into which the felt partially sinks there will be a certain amount of drag and polishing action, combined with the removal of material not only from the flat portions but to some extent from the depressions also. There will be a general rounding off of the irregularities, but, as is confirmed by practice, the irregularities cannot be eliminated by the use of a soft polisher, although a small amount of original greyness is rendered less conspicuous to ordinary vision, and a false appearance of complete uniformity and polish may be produced more quickly by means of a soft polisher. Suppose a harder type of polisher such as a pitch polisher is used. When using the same polishing medium, such as rouge, the amount of drag and polishing action over the flat portions will be greater than when using cloth, and, further, as the

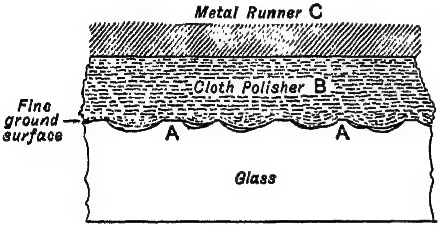


FIG. 12.

pitch will not sink into the depressions to the same extent, the rounding and particularly the removal of material from the depressed portions will be less.

When using a hard polisher a uniform polished surface free from all greyness is obtained by the removal of the surface material stage by stage until the surface is reduced below the level of the deepest conchoidal depression.[1]

Under the action of the polisher, according to the flow theory, the disturbance of the surface molecules is such that they are able to rearrange themselves or flow with the consequent formation of a polished amorphous layer. Minute aggregations of the rouge or other medium plough away the surface layer, and it is conceivable that there may be also some swaging action, material being removed from the higher portions and welded upon the adjacent depressed areas. As the amorphous material is removed in this way the underlying molecules are acted upon by the polisher, and the process is repeated layer by layer as indicated in *Fig.* 13.

If rouge is employed in the last stage it is generally possible by special illumination of the surface to detect an open network of these fine grooves; but if no medium other than a very fine film of water is used for the final polishing operation, the presence of grooves will hardly be observable. When a surface of this kind is etched with hydrofluoric acid a network of grooves will reappear,[2] and it has

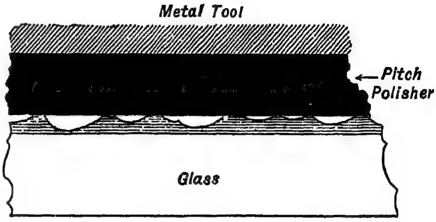


FIG. 13.

been assumed that just as in the case of metals the original grooves have been bridged over by the amorphous layer and are uncovered when the surface layer is dissolved away. Numerous experiments, however, seem to indicate that the cohesion of the amorphous silicates is too small to permit of the bridging of the finest surface cracks that can be produced.[3] That grooves produced during the rouge-polishing stage become filled up during the final water-polishing stage may, however, be accepted. Thus it is probable that the groove A in the surface layer becomes filled up possibly in stages at each stroke in the manner indicated in *Fig.* 14. On the assumption that the expenditure of energy upon a substance tends in general to reduce its chemical stability, it is to be expected that the material filling the grooves would be rapidly acted upon and that the grooves produced by

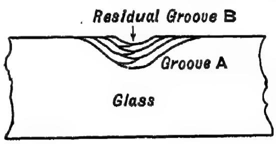


FIG. 14.

etching are really reproductions of the original grooves. The course of the original grooves will in any case be indicated by a very fine groove B, which would be extended by the acid to form a deeper groove occupying the place of the original one.

The various stages in the grinding and polishing of a glass surface may be summarised briefly as follows:

1. The removal of material by the breaking away of splinters, the size of which is reduced in stages by the use of finer grades of abrasive.

[1] "Polishing of Glass Surfaces," Lord Rayleigh, *Proc. Opt. Convention*, 1905, i. 75.

[2] "Interference Bands and their Applications," Lord Rayleigh, Royal Inst., March 24, 1893.

[3] "Some Notes on Glass Grinding and Polishing," J. W. French, *Trans. Opt. Soc.*, Nov. 1916, xvii. No. 2.

2. The production of an amorphous or surface flow layer and the gradual removal of these layers by grooving as distinct from the splintering of the first stage, the removal of material being effected by means of a very fine abrasive or polishing material such as minute aggregates of particles of rouge.

3. The elimination of the grooves produced in stage 2 by the use of a continuous medium such as a film of water in place of the discontinuous medium such as rouge, there being during this stage the maximum production of surface flow and practically no removal of the surface layer material and no splintering action whatever.

§ (8) Polishing Materials.—An abrasive to be effective must possess several well-defined characteristics. The grains, which must be hard, should have an irregular form presenting many strong edges or rounded points that will transmit the impact forces to the glass to be abraded. When the grains break down the fragments should be of the original form in order that the action on a finer scale may be continued. An abrasive that breaks down into lamellar fragments is said to lose its cut. Diamond, splinters of which are illustrated in *Fig.* 15, is the most effective of abrasives, but owing to its cost it can only be used for special operations such as the slitting of glass where the quantities are small.

Carborundum (SiC) is a compound of carbon

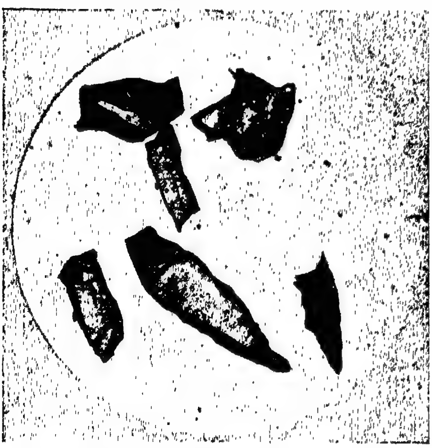


Fig. 15.—Diamond Splinters. Magnification 15 diameters.

and silica, resulting from the fusion of carbonaceous materials such as coke or charcoal with sand in the electric furnace at a temperature of about 2000° C. Its hardness on Mohs scale is about 9, diamond being 10. From *Fig.* 16 it will be seen that the grains are of the desired irregular shape, which is retained as they break down.

Other important abrasives are obtained by the combination of alumina and silica in the electric furnace, such, for example, as corundum, alundum, and aloxite. They are

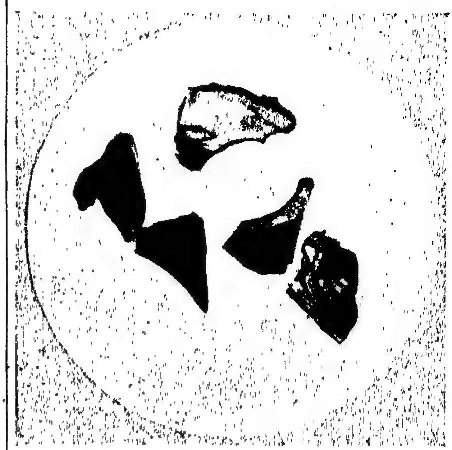


Fig. 16.—Carborundum. Magnification 40 diameters.

most commonly employed in the form of grinding wheels. As loose abrasives they are not so effective as carborundum.

Emery is a natural form of artificial corundum, being a silicate of alumina containing, however, oxides of iron and other impurities irregularly distributed. It breaks down more readily than carborundum and loses its cut. The loss of cut is only temporary, however, as after washing the material can be used as a finer grade of emery.

Sand is sometimes used for rough abrasion, more particularly in establishments where no facilities exist for the washing and recovery of the more expensive types of abrasive. Its grains are frequently rounded and water-worn as indicated in *Fig.* 17, and it readily breaks down and loses its cut.

Separation of an abrasive into the various grades of fineness is generally done by a process of settling and levigation. Thus three-minute emery is the material obtained from the liquid that is decanted after a settling period of three minutes. For carborundum it is necessary to use sieves of various finenesses in conjunction with settling, except in the finest stages, for which sufficiently fine sieves are unprocurable.[1]

A comparison of the abrasive powers of carborundum, emery, and sand of approximately equal size of grain is obtainable from *Fig.* 18, which also shows how in the case of each abrasive the rate of abrasion is directly proportional to the pressure when fresh abrasive is supplied continuously.

If the abrasive is not frequently renewed,

[1] "The Grading of Carborundum for Optical Purposes," J. W. French, *Trans. Opt. Soc.*, Oct. 1917.

the rate of grinding would not increase regularly with the load owing to the loss of

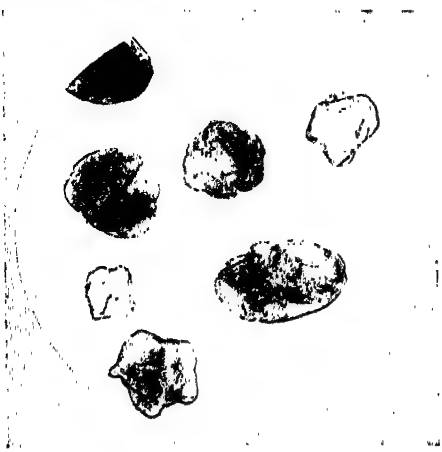


FIG. 17.—Sand. Magnification 40 diameters.

cut and a certain clogging action, especially at the higher pressures. This is illustrated in *Fig.* 19, from which it will be seen that when using No. 3 carborundum having an average grain diameter of 0·1 mm. the rate of abrasion is directly proportional to the load when fresh

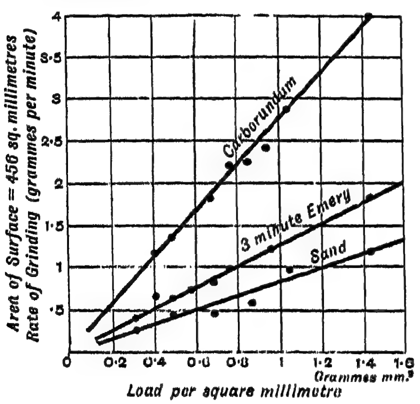
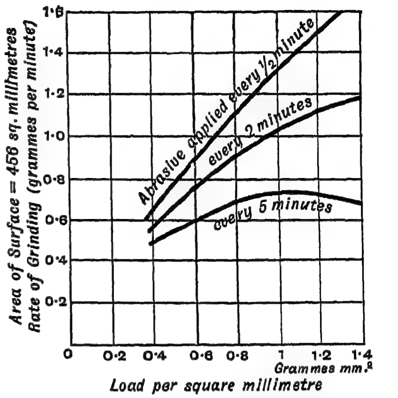


FIG. 18.

material is applied every half minute, whereas there is actually a reduction of the rate at high loads when the intervals between the applications are of five minutes' duration.

The series of curves in *Fig.* 20 shows how the rate of abrasion varies with the size of the grains and the load on the tool. Fresh abrasive was applied every half minute, and it will be seen that for the coarsest No. 1 carborundum having a grain diameter of approximately 0·2 mm. the abrasion is directly proportional to the load. In the case of the carborundum grades 3, 4, and 5 the curve bends at the higher loads suggesting the need for more frequent renewal of the abrasive, due possibly to clogging arising from an admixture of glass powder in these intermediate grades.

A similar series of abrasive curves for 3 minute, 15 minute, and 40 minute emery is



No. 3 Grade (washed) Carborundum (0·1 mm. diam.)

FIG. 19.

illustrated in *Fig.* 21. Between the 3 and 15 minute curves there is a considerable interval which corresponds, however, with the grain dimensions, which are 0·13 mm. and 0·01 mm. respectively.

From the various diagrams it will be evident

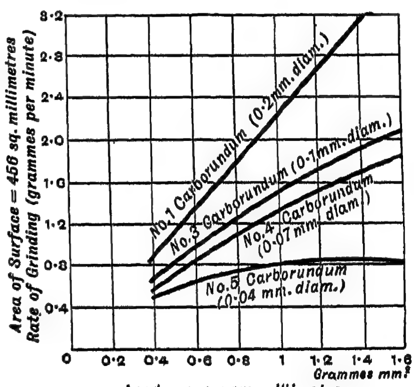


FIG. 20.

that the rate of abrasion of a particular type of glass depends upon:

The nature of the abrasive.

Its grade of fineness.

The load upon the tool.

The relative speed and to some extent the frequency of renewal.

The material of the tool itself, which has some bearing on the rate of abrasion, is

practically determined by the type of abrasive, the materials of widest application being fine-grained cast iron and brass.

Rates of abrasion afford a general but not definite idea of the quality of a ground surface, because the same weight of material may be removed by the production of large shallow splinters or by smaller and correspondingly deeper ones. In the grinding process it is sought to produce a surface of uniform texture free from isolated deep pits which often determine the thickness that must be finally removed. A record of the texture may be obtained by observing the reflecting power of the ground surface.[1] If the surface is viewed so obliquely that the irregularities are foreshortened to such an extent that even the longest red rays are reflected, a perfect white image of, say, an incandescent filament may be seen reflected from the surface. The image

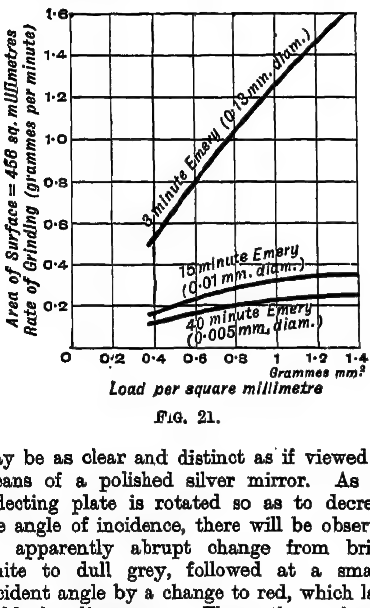


FIG. 21.

may be as clear and distinct as if viewed by means of a polished silver mirror. As the reflecting plate is rotated so as to decrease the angle of incidence, there will be observed an apparently abrupt change from bright white to dull grey, followed at a smaller incident angle by a change to red, which later suddenly disappears. These three changes take place sufficiently abruptly to provide a general record of the surface. Thus in the case of a piece of hard crown glass, ground with 3 minute emery, the respective angles of incidence, the readings of which can be repeated to within half a degree, were 80°, 78°, and 75°. For similar glass ground with an abrasive wheel, the corresponding figures were 66°, 61°, and 46°, thus indicating a much finer texture. In *Fig.* 22, which gives a comparison of the surfaces produced by a series of carborundum abrasives, abscissae represent the size of the respective grains, and ordinates the cosecants of the angle between the surface and the line of sight. This is equivalent to the projection of the texture in a plane normal to the line of sight, thus enabling the irregularities to be compared directly with the lengths of the waves regularly reflected. As is to be ex-

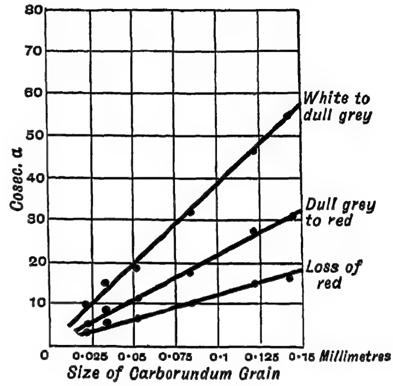


FIG. 22.

pected from the previous abrasion diagrams the curves are straight lines, the projected dimensions of the texture being directly proportional to the size of the grain when the other conditions such as load and speed are maintained constant throughout.

Fig. 23 shows the corresponding results obtained when using the three grades of emery commonly employed, namely 3 minute, 15 minute, and 40 minute emery. If the conditions can be controlled with sufficient accuracy, similar straight line curves may be obtained for most abrasives, but in the case of such abrasives as powdered glass which break down readily and lose their cut, the conditions cannot

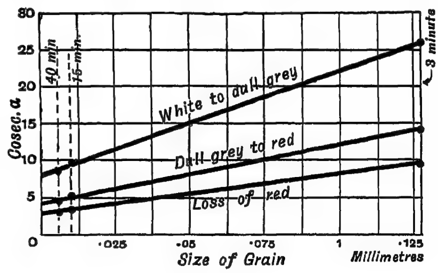


FIG. 23.

be easily controlled and the curves generally fall away towards the coarser grades.

Smoothing operations in the case of a single piece may be regarded as a continuation of the rougher grinding processes, their purpose being to reduce as much as possible the amount of material that has to be removed in the polishing process, but when a number of pieces after

[1] "More Notes on Glass Grinding and Polishing," J. W. French, *Trans. Opt. Soc.*, 1917 (Jan.), vol. xviii.

being formed are mounted in one block so that they may be polished together, the smoothing process preparatory to polishing is necessary in order to reduce any irregularity of the surface levels due to slight errors in the laying down of the pieces.

Whether the surface desired is plane or curved the final form of the piece must be produced before the polishing operation is commenced. Any important alteration of the form is impracticable during the polishing process, the functions of which are the formation of the brilliant amorphous light reflecting surface layer, and the production of a true figure, that is, the correction of minute errors of form not exceeding one or two wave-lengths over the whole surface. Greater errors of form must be corrected by a repetition of the smoothing or fine grinding operation.

Almost any substance in a fine enough state of division, and provided its grains are not lamellar in form or soluble in the liquid employed or liable to weld upon the surface of the glass, may be used as a polishing medium. Thus glass can be polished readily with fine charcoal, but hardly at all with graphite.

But in practice the choice may be limited to a very few substances, the principal of which is rouge, that is oxide of iron (Fe_2O_3). Many substances are slow in their polishing action. Putty powder, that is tin oxide (SnO_2), which at one time was extensively used, is now excluded for reasons of health.

Manganese dioxide (MnO_2), although an excellent medium, is very black and is difficult to remove from the hands and clothing; from the point of view of general cleanliness its use is often avoided. Other media again cannot be obtained in a consistently uniform condition, and although the variety of substances is great, there are really few that have all the advantages of rouge, which is so extensively used, except for the very cheapest kinds of optical work. Comparisons of the polishing media can only be made if the conditions are carefully standardised, and particularly if the texture of the original smoothed surface is the same in all cases. The rate of polishing, so far as polishing is determined by the removal of material until the bottoms of the deepest depressions are reached, depends upon the following:

The original state of the smoothed surface.

The character of the polishing-tool surface, which may be, for example, of cloth, pitch, wax, or paper.

The polishing medium.

The lubricant.

The load.

The relative speed of the tool and the surface operated upon.

The typical chart, *Fig.* 24, shows how the time of polishing and the rate of removal of glass are affected by the load on the tool. In this particular instance the rate of removal of material varied directly with the pressure, the tool being covered with a mixture of 95 per cent of beeswax and resin and 5 per cent pitch.

It is not possible to make any definite comparison of the numerous substances that may be used as polishing media, because the results are greatly influenced by the conditions. Substances that are only moderately good when a pitch polisher is used may be much more effective when the polisher is of a different type, such as cloth.

To obtain consistent results it is also very necessary to control the conditions, the most important being the original state of the smoothed surface to be operated upon.

Table I. shows the times of polishing when using a variety of typical polishing materials under the particular conditions specified. In all cases the original surface was smoothed with fine carborundum having an average grain diameter of about 1/200 mm., the reflection values of the surface being 80° grey, 73·5° red, and 64° loss of red. Pitch polishers were used throughout.

TABLE I

Machine type, reciprocating arm.
Revs. of spindle, 78 r.p.m.
Speed of arm, 190 strokes per minute.
Length of stroke, 1·25″.
Diam. of worked surface, 3″.
Diameter of polisher, 3″.
Load on polisher, ½ lb. per sq. inch.

Medium.	Polishing Time.	Quality.
	Hours.	
Precipitated rouge	3	Good polish.
Commercial rouge	4	
Glassite	4·1	
Very fine carborundum	8	Fair.
Putty powder (SnO_2)	11	Surface cut and not good.
Precipitated silica (SiO_2)	12	Surface not good.
Precipitated chromium oxide (Cr_2O_3)	14	
Precipitated alumina (Al_2O_3)	14	
Precipitated ferrous carbonate	16	Very slightly grey.
Precipitated hydrous MnO_2	22	

Putty powder and the various precipitated media are better suited to cloth polishers than to pitch polishers, but the

superiority of the first three media is still marked when used on cloth.

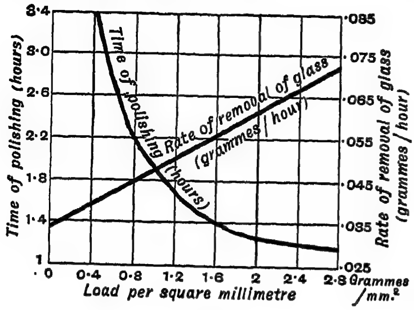


FIG. 24.

Very many substances have been used as a covering layer for the polishing tool to hold the polishing medium.

Huyghens [1] polished his glasses upon the metal tool itself, using a specially prepared mixture of Tripoli as a medium, which was reduced to a firmer and finer state by wiping away the marginal portions from time to time.

To Sir Isaac Newton [2] is attributable at least the suggestion that a pitch layer might be used for the polishing operation. In his *Opticks* he states that: "An object glass of a fourteen foot telescope made by one of our London artificers I once mended considerably by grinding it on pitch with putty and leaning very easily on it in the grinding, lest the putty should scratch it. Whether this way may not do well enough for polishing these reflecting glasses I have not yet tried."

It is quite clear from a previous paragraph [3] that Newton actually used pitch for both the grinding and the polishing of metal reflectors, for in the same work he states: "Then I put fresh putty upon the pitch and ground it again till it had done making a noise, and afterwards ground the object metal upon it as before, and this work I repeated until the metal was polished, grinding it the last time with all my strength together for a good while, and frequently breathing upon the pitch to keep it moist without laying on any more fresh putty." The application of water alone in the final stage is of particular interest as being an important detail of present-day practice.

Father Cherubin [4] lined his polishing tools with a variety of materials of fine and uniform texture. More particularly he refers to the use of very fine thin leather, fine English fustian, fine Holland or any fine linen, silk taffety or satin. He describes at great length the process of lining the tools with paper and the method of removing little lumps or irregularities, but the previous use of paper is attributable to Antheaulme,[5] who used the grinding tool itself when lined with paper as the polishing tool, the medium employed being Venetian Tripoli.

At the present day pitch, wax compounds and cloth are the materials most commonly used. Comparisons of a variety of materials when using a particular polishing medium can be obtained from Table II., which shows

TABLE II

SURFACE LAYERS OF POLISHING TOOLS

Material worked, glass—hard crown.
Surface smoothed with fine carborundum, diam. of grain about 1/200 mm.
Spindle speed, 124 to 130 r.p.m.
Arm speed, 100.
Stroke, 1·25″.
Polishing medium, rouge and water.
Load on polisher, ½ lb. per sq. inch.

Polisher.	Time of Polishing.	Remarks.
	Hours.	
90% pitch with 10% beeswax and resin	1·5	Good surface.
Very hard pitch	1·75	Fair surface, pitch hardness 0·5 at 37° C.
Beeswax and resin with 10% pitch	1·75	Good surface.
Pitch 70% with 30% rubber compound	2	" "
Pitch 70% with 30% rouge .	2	Fair surface.
Soft pitch	2·75	Good surface, pitch hardness 16 at 38° C.
Ebonite with 10% pitch . .	3·5	Polished, but cut, contact obtained with difficulty.
New cloth	4·5	Good surface.
Well-singed cloth . . .	6	Poor surface, cut.
Cork	6	Fair surface, slightly cut owing to bad contact.
Wood, deal	6	Cuts, owing to difficulty of obtaining good contact.
Brown paper, impregnated with beeswax and resin	6	Fair surface.
Very thick felt	12	Surface grey and mottled.

[1] C. Hugenii, *Opera Reliqua*, ii. 218, and Smith's *Optics*.
[2] Newton, *Opticks*, p. 78.
[3] *Ibid.* p. 77.
[4] Le Père Cherubin, *La Dioptrique oculaire*, 1671, part iii. chap. ii.
[5] Montucla, *Histoire des mathématiques*, vol. iii. part v. book ii. p. 498.

the time required to polish a piece of crown glass smoothed in all cases to the same standard. For the purposes of comparison comparatively slow speeds and loads were employed. From the previous diagrams it will be understood that the time of polishing would be reduced by increasing these factors, and the relative positions of the various materials might be slightly modified by the use of another polishing medium such as putty powder.

Since the time of Newton, over two hundred years ago, hand-polishing methods as practised to some extent in most high-class workshops have remained unaltered, except possibly in

production of an optically flat surface; if two are ground together the respective surfaces will become convex and correspondingly concave, the tendency in practice being for the upper tool to become concave. If one of these tools, say the convex No. 1, is then ground with No. 3 tool, the latter will become correspondingly concave, more or less like No. 2. If then the two concave tools Nos. 2 and 3 are ground together, their concavities

FIG. 25.—Hand Working.

so far as the materials employed are more uniform in quality.

§ (9) POLISHING TOOLS.—Some idea of the essentials of the actual flat or prism surface polishing operations may be obtained from *Fig.* 25. The pedestal (1), which must be rigid, has at its upper end a standard nose-piece upon which the tools may be screwed. Under the work bench there will be observed two flat tools (2) and (3) of close-grained and well-annealed cast iron, and on the bench another (5) lying face downwards with a wooden operating knob screwed into the boss. On the pedestal is mounted the tool or runner (4) as it is called, which has a uniform flat surface layer of pitch, and on the bench there is a similar runner (11), the pitch surface of which is exposed to view.

Three plane tools are necessary for the

will both be reduced. By grinding No. 1 on No. 2, then No. 1 on No. 3, and No. 2 on No. 3 in this way, and repeating the sequence of operations as long as may be necessary, all three tools become optically flat and may be kept in this condition by an occasional repetition of the process.

Sir Joseph Whitworth's name is generally associated with this method, which he applied to the production of standard surface tables, but although the principle may not previously have been clearly expressed, the method appears to have been known to the earlier opticians.

The work is ground or smoothed upon one of the metal tools, fine grades of carborundum or emery being used as the abrasive medium. It is essential that the surface should be ground and smoothed to the desired curvature

or flatness, as the time of polishing depends upon the perfection of the smoothing operations. Removal of material during the polishing process occurs very slowly, and it is then impracticable to effect an alteration of the shape other than a change of the figure involving a removal of material to a depth of a few wave-lengths.

For the polishing process the surface of the metal runner, which itself need only be approximately true, is covered with a layer of pitch about 12 mm. thick, the runner being heated gently to ensure good adherence. The pitch surface is then moulded by one of the flat tools that has been heated just sufficiently to soften the pitch upon which it is pressed. When the pitch is cold, the whole surface is divided into small squares by deep grooves which may be cut more cleanly under water when the pitch is of a brittle type. Over the pitch surface there is then stretched a piece of open texture muslin which is squeezed into the pitch by means of the hot flat tool, and is then peeled off, leaving its network impression on the surface.

While the deep grooves help to preserve the flatness by breaking the continuity of the layer and thus preventing the centre parts from being squeezed towards the periphery, their principal function as well as that of the fine network is to destroy the suction that would hinder the free movement of the work over the moist pitch surface.

The grooves also serve to retain any excess material that otherwise might collect and produce sleeks, that is, minute furrows on the surface of the work, or even cuts.

After the pitch surface has been prepared in the manner described, it is rubbed with one of the optically flat metal tools until it also is optically flat. Rouge and water are commonly used as the working medium, not only in the preparation of the pitch surface but also in the operation of polishing the work.

By means of the soft brush (7) fine, well-levigated rouge and water from the pot (9) are laid in streaks on the pitch surface of the tool (4), and the harder brush (6) is employed to spread the medium uniformly over the whole surface upon which the part to be worked is laid, the sponge (8) being used throughout the operations to wash away excess material from around the edge of the tool.

The operator, whose attitude is illustrated in *Fig.* 25, controls the work around its periphery by means of the fingers and thumbs, and, while gently pressing it into contact with the pitch surface, he moves it to and fro. At frequent intervals, after a few such repeated strokes, the work is given a wide-sweeping movement over the pitch surface and is occasionally rotated. At intervals the operator also steps round the pedestal, or, alternatively, the lower tool may be slowly rotated.

The purpose of these movements is to avoid any regular repetition of strokes that would tend to local wear of the tool or work, and the production of irregular surfaces. If the to-and-fro movements in one direction were repeated for too long a time, a broad depression would be formed on the tool surface, the optical flatness of which would accordingly be destroyed.

From time to time the pitch surface is reworked or formed by means of the metal tool (5), the flatness of which is preserved by occasional working with the tools 2 and 3. Towards the end of the process water only is used as a polishing medium and the operation is continued until the water is almost entirely dried up, which is evidenced by a characteristic squeaking noise. In this way the greatest possible viscous flow of the surface molecules is obtained owing to the close contact between the work and the polishing tool.

Surface defects such as sleeks are detected by examining the surface with a low-power lens, the necessary illumination being obtained from the lamp (10). Defects of flatness are detected by means of the test plate (12), which is placed on the surface of the work lying upon a black cloth. It is essential that the temperature of the parts should be allowed to equalise before attempting to form a definite decision as to the character of the surface under test, and before using the test plate the parts should be thoroughly cleaned with a linen cloth or selvyt (13), and also freed from dust by means of the soft brush (14).

At night it is convenient to use mercury-vapour lamplight, which is approximately homogeneous and produces black interference rings that may be very readily observed.

Much skill is necessary to obtain the correct figure or form of the polished surface within the limits essential to the production of well-defined images, and each operator has his own particular method of controlling the figure, which indeed to some extent depends upon individual characteristics such as the temperature of his hands. Two methods are frequently employed. In handwork, with the tool below the glass which has the smaller area, if the surface becomes concave the stroke should be reduced within the limits of the polisher. If the surface becomes convex the stroke should be widened until the defect is corrected. On machine work, with the tool having now the smaller area above the work, these operations should be reversed, a wide stroke, and especially one that overlaps the polishing tool, being employed to reduce any concavity of the surface of the work.

In the case of large surfaces, such as are usual in machine multiple work, the position of the polishing tool may be so altered that it acts more around the periphery or over the centre, as may be desired to correct the figure.

The second method consists in ringing the tool, that is, in broadening the furrows or scraping the tool surfaces, and thus reducing the effective polishing area at the centre or towards the edge, according to the requirements.

Thus, suppose a block of concave lenses is too shallow. To correct the figure it is necessary to remove material from the centre. This may be done by lengthening the stroke of the block so that at the ends of the travel it overlaps the polishing tool, with the result that for a portion of the time the outer parts, as compared with the centre, are not acted upon and are not reduced to the same extent; or the outer portion of the polishing tool may be reduced by scraping, or the effective surface reduced by ringing, so as to increase the relative effect of the centre parts.

A polisher that has been used for a considerable time often becomes glazed and loses its effect, so far as the removal of glass is concerned, but not as regards the actual polishing action if the contact is good. It has been suggested that the glaze which may be readily scraped away is a more or less continuous layer of glass.[1]

Particular results are also obtained by a proper selection of the polishing layer.

In the case of pitch, which should be a good quality of Burgundy pitch, the hardness must be varied by more or less prolonged boiling to suit the temperature, and to some extent the nature, of the glass and work.

Thus, unless the workshop temperature can be kept constant, which is not easy so far as the reduction of the maximum summer temperature is concerned, a hard, well-boiled pitch must be used in summer and a softer pitch in winter.

Some operators prefer the use of wax instead of pitch, and particularly for the polishing of curved work.

From *Fig.* 26 it will be seen that there are characteristic differences attributable largely to differences in the viscosities of the materials. Abscissae represent temperatures and ordinates the rates of penetration of a steel disc under a constant load. Whereas the pitch yields, even at the low temperatures, very little change occurs in the beeswax-resin composition until a temperature of about 27° C. is reached, when the viscosity rapidly changes. As the normal temperature of working rarely exceeds 22° C., it will be evident that wax layers are not so susceptible to fluctuations of temperature as pitch layers. Wax therefore retains its shape better than pitch, but its form is not so easily manipulated when a modification of

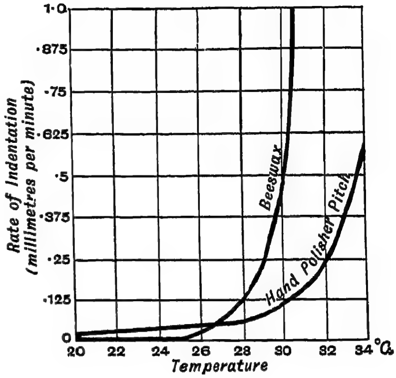


FIG. 26.

the figure is desired, and the difficulty of obtaining good contact is greater.

§ (10) POLISHING PROCESSES.—Only the very highest qualities of optical work or small quantities of individual parts are made by skilled hand methods. For the production of parts in large quantities considerations of cost make the use of machinery essential. Although highly specialised machinery has been evolved for the manufacture of particular products, such as spectacle lenses,[2] there are certain well-defined methods and types of machinery, a description of which alone will suffice to indicate the fundamental principles.

The machine processes may be divided, as in the case of hand work, into the three groups:

(*a*) Forming or roughing.
(*b*) Smoothing.
(*c*) Polishing.

Frequently the raw glass is supplied in the form of plates which are cut to the approximate shape by means of saws, as illustrated in *Fig.* 27. These saws are thin sheet-metal discs of a soft character. Iron armature stampings are very suitable for the purpose. The edge of the saw is notched and charged with diamond dust mixed with oil to the consistency of a fine paste. In Germany it is common practice to cut deep radial peripheral slots about a millimetre broad and to fill these with lead, which holds the diamond dust more effectively.

The work, which may be a pile of plates held within the jaws of an adjustable holder, is pulled by a weight against the cutting edge, the movement being controlled by the withdrawal of a screwed abutment.

[1] "The Surface Layer of an Optical Polishing Tool," J. W. French, *Proc. Opt. Soc.*, 1920, xxv. No. 3.

[2] *Schule der Optik*, Gleichen und Klein, 1914, practical section by Klein.

Turpentine as a lubricant gives the best results, but it has the disadvantage of being costly. Petrol is very suitable. Water with a small admixture of soda to prevent rusting of steel parts, although less effective, is commonly used.

An average peripheral speed of the saw is about 1200 feet per minute.

Labour and expense in cutting may be saved by the use of glass that has been moulded to approximately the correct shape and size, enough excess material only being allowed to ensure that when all the irregularities of the surface have been ground away the sizes will not be too small.

When the quantities involved are not large, the actual forming of the work to the required linear dimensions and shape within usually

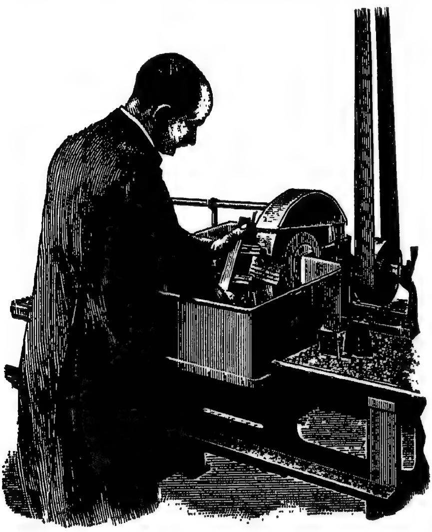


FIG. 27.—Disc Sawing Machine.

an angular limit of $\pm$ three minutes is sometimes done by hand, as indicated in *Fig.* 28, the angles being tested by means of simple adjustable gauges[1] or, in the case of lenses, by disc gauges turned to the appropriate radius. For the roughest grinding, sand or coarse carborundum or emery may be used, the sand being cheapest but slowest in action. If a plant for the washing and grading of the abrasive is installed,[2] the use of a single type of abrasive such as carborundum, a coarse and cheap grade of which only need be purchased, is both economical and convenient.

When two abrasives such as carborundum and emery are employed for the coarser and finer grinding processes respectively, great care must be exercised to avoid contamination of the emery with the harder carborundum. The large cast-iron grinding disc in *Fig.* 28, which

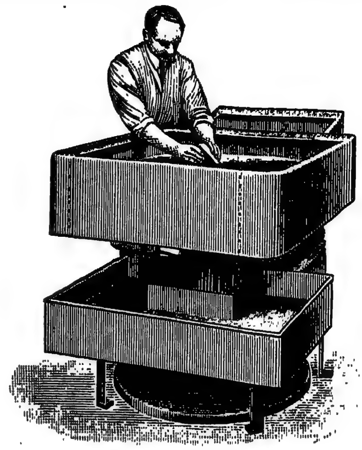


FIG. 28.—Rough Grinding Machine.

runs at a speed of about 250 revs. per minute, serves the double purpose of breaking down the coarser abrasive preparatory to fine grading and of forming the heavier work.

In the case of multiple work the cost of labour generally necessitates the adoption of grinding

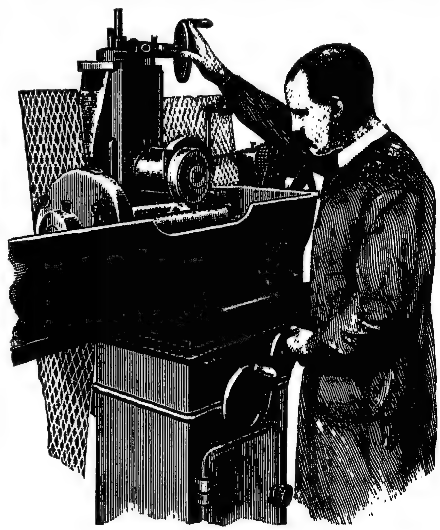


FIG. 29.—Glass Milling Machine.

machinery, one type of which is illustrated in *Fig.* 29. A number of parts are mounted upon an adjustable holder on the table of the machine, the holder being so arranged that it can be accurately angled in accordance with the indications of a scale. The table

[1] *Handbuch der praktischen Optik*, by Halle, 1913.
[2] "The Grading of Carborundum for Optical Purposes," *Trans. Opt. Soc.*, Oct. 1917.

runs the work longitudinally under the grinding wheel, and at the end of each stroke its motion is automatically reversed, the grinding being done during both strokes. At the end of each stroke the table is also fed transversely, automatically or by hand. Either a heavy cut of about 1 mm. with a slow table speed or a light cut of $\frac{1}{50}$ mm. with a fast cutting speed of 10 feet per minute may be used, but generally it is desirable to avoid the heavy forces incidental to a heavy cut.

The grinding wheel, which is fed down by hand as required, is a fine-grade carborundum type, such as 220 J, running at a peripheral speed of about 500 feet per minute.

An ample supply of lubricant, such as water, is essential to avoid all danger of the formation of minute heat cracks.

Milling cutters in the form of cylinders of copper, the surfaces of which are grooved longitudinally and charged with diamond dust, have been used by some of the more important German opticians, but such cutters now appear to have been generally abandoned, except for certain minor operations, in favour of carborundum wheels.

Glass milling, as compared with surface grinding, has the disadvantage of being comparatively slow and therefore relatively costly. This will be evident when it is considered that there is only line contact between the wheel and the surface of the work, which for the great part of the time is therefore not being acted upon, whereas in the case of surface grinding, material is being removed continuously from every part of the surface.

Surface grinding is usually done by means of a cast-iron disc with loose abrasive, and the parts to be operated upon are mounted in accurate multiple jigs, the expense of which can only be contemplated when the number of pieces to be manufactured is large.

As the work must be formed practically within the final inspection limits, it is necessary that the machine should be capable of working to an angular limit of about 2 minutes and a dimensional limit of about $\frac{1}{10}$ mm. to $\frac{1}{40}$ mm.

After the individual parts have been formed to the exact shape they are submitted to the smoothing processes preparatory to being polished. For this purpose the faces to be polished are placed in contact with an optically worked tool, *Fig.* 30, the surface of which has been thoroughly cleaned and slightly oiled. Care must be taken to ensure that the layer of oil between the surfaces of the work and the tool is not too thick, as otherwise a very slight pressure on one end of the piece may suffice to introduce an angular error of a minute. With a little practice such errors, even in the case of faces of a few centimetres length, may be avoided.

Over the tool there is then placed a circular framework, into which is poured a layer of plaster of Paris of about 3 mm. thickness. After the plaster has solidified the frame is filled with a special cement whose volume remains practically constant after solidification.

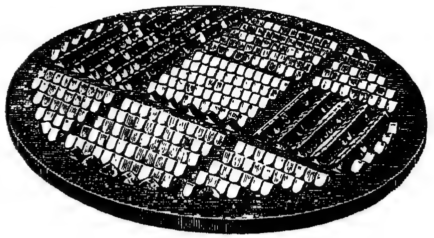


FIG. 30.—Block of Prism.

Plaster of Paris and cements that contain free lime have the disadvantage that, owing to crystalline changes in the solid condition, the volume after solidification increases for many months, the growth being comparatively rapid during the first few weeks.

In such a case, owing to this expansion, the smoothed surface of a cemented block of prisms tends to warp to an extent that may unduly prolong the polishing time, or even necessitate re-smoothing.

Plaster of Paris mixtures are suitable for blocks of 25 cm. diameter, and can be used for blocks of 50 cm. diameter if there is no delay between the smoothing and polishing processes and if the latter is completed rapidly and in one operation. For large blocks of 1 metre diameter, such as is illustrated in *Fig.* 31, special cements are essential.

The frame with the solidified cement containing the prisms is stripped from the oiled

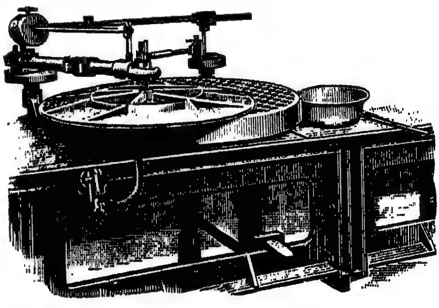


FIG. 31.—Smoothing or Polishing Machine for Blocks 1 metre diameter.

surface of the plate, and when turned over the thin layer of plaster of Paris can be removed, leaving the surfaces of the prisms projecting from the layer of cement, the face of which is cleaned and varnished to make it thoroughly waterproof and to prevent fragments from breaking loose and interfering with the work

of polishing. The thin layer of plaster of Paris has no appreciable ill-effect, as its thickness and time of action are both small.

For the actual smoothing operation a machine of the general type, originally devised by Lord Rosse,[1] indicated in *Fig.* 31, is employed, but although the principle involved is generally the same in all cases, the arrangements vary considerably.

In a machine tool for the working of metals or the forming of glass, such as is indicated in *Fig.* 29, the plane in which the work moves and the axis about which the tool rotates are both definitely fixed. The accuracy of the work is accordingly dependent upon the accuracy of the machine.

As the accuracy requisite for the polishing of an optically good surface is nearly ten times as great as is obtainable with the best type of machine tool, it is necessary for the manufacture of optical work to adopt the floating tool principle previously referred to in connection with the preparation of flat grinding tools.

In *Fig.* 31 it will be seen that the flat grinding tool rests freely upon the surface of the work, the diameter of which is about 20 per cent greater than that of the tool,[2] and that it is moved by a central pin resting in the hollow socket of the tool. For the actual movement of the tool various mechanisms have been introduced, but that indicated in the illustration is the most common.

All the arrangements[3] are particularly designed to sweep the tool over the surface in a continually and widely varying path that only repeats itself after a great number of strokes, and further, to provide simple adjustments whereby the action of the tool may be distributed more or less over the central or outer zones of the work for the purpose of regulating the form or figure.

Behind the machine there are situated two vertical shafts each carrying at its head a crank-pin the distance of which from the centre of rotation is adjustable. One crank drives the end of the main driving arm through the intermediary of a gimbal connection that enables the arm to be raised when necessary quite clear of the tool. The other crank similarly drives a radius bar coupled to the driving arm. As the throws of the crank are independently variable and as the point of connection of the radius bar with the driving arm is adjustable, a large variation in the path of the tool is obtainable. It is important that the revolutions of the crank spindles should not be a whole multiple of the work spindle revolutions. A hunting tooth should be introduced in the gearing, otherwise a portion of the surface will take longer to smooth or polish, thus increasing the time of the operation. By adjusting the length of the radius bar the tool may be made to act more over the periphery of the work, and the length of the stroke may be altered by a variation of the amount of throw of both cranks. These adjustments are necessary for the control of the form or figure of the work.

To determine by calculation the distribution of the grinding or polishing action over the surface of the work is very difficult and laborious. As was previously stated, the action is practically directly proportional to the relative speed of the tool and the work. When the tool is concentric with the rotating work, both rotate together, and there is no relative movement or grinding action unless a partial brake is applied to the tool. As the tool passes from the centre towards the periphery its velocity changes, and there is then relative motion of the tool and work and consequently abrasion. This rotation is always in the same sense as that of the work, but it necessarily varies throughout the cycle, although the momentum of a heavy tool acquired when in the concentric position tends to make the rotation more uniform.

The action at any particular point depends upon the time the part is acted upon by the tool, and this varies with the relative positions of the tool and the work, as the stroke is generally such that the tool sometimes overlaps the work at continually varying positions and by varying amounts.

The action is also proportional to the load, and this is not constant as the area of the tool in contact with the work varies in the overlapping positions.

From these considerations it will be evident that the problem of determining the distribution of the action even under the simplest conditions is a very complex one.

The object of the smoothing process is to remove irregularities of the total surface arising principally from small errors in the laying down or assembling of the individual parts and minor distortions of the mass, and to remove any accidental small holes or cuts introduced by the coarser abrasive during forming. If the initial work has been well done the amount of material to be removed is small, but as the rate of removal of material during the smoothing operation, when fine emery or carborundum is used as an abrasive, is much more rapid than during the polishing stage when rouge is employed, the importance of good smoothing will be evident.

The polishing process is practically a continuation of the smoothing, the same methods and type of machine being employed; but a polishing medium such as rouge is used in the first instance instead of an abrasive, and a pitch, wax, or cloth polisher instead

[1] Lord Rosse's *Telescopes*, 1884, and *Phil. Trans.*, 1840.
[2] *Deutsche mech. Zeit.*, 1909, ix. 81.
[3] *Handbuch der praktischen Optik*, Halle, 1913.

of a metal tool. As in the hand-polishing process previously described, water only is used as a polishing medium in the final stages of the operations.

§ (11) POLISHING MACHINE.—Another type of machine capable of polishing a block about 1·5 metres in diameter is indicated in *Fig.* 32,

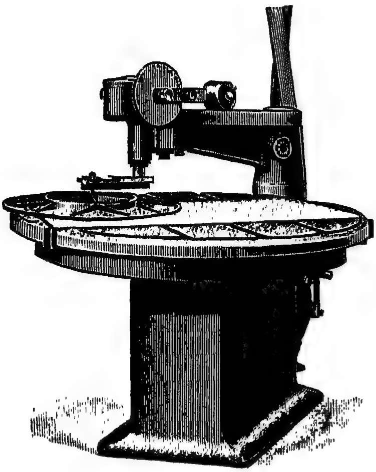


FIG. 32.—Smoothing or Polishing Machine for Blocks 1·5 metres diameter.

and the working of a concave block of large diameter lenses is illustrated in *Fig.* 33.

Two methods may be employed for the

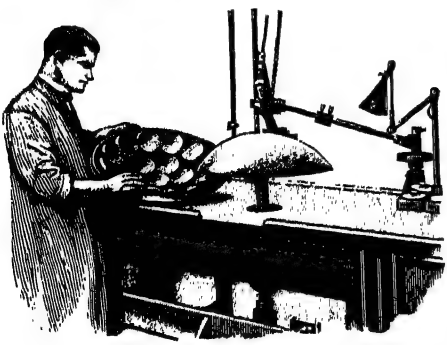


FIG. 33.—Block of Concave Lenses.

production of curved work such as the elements of an objective or an eyepiece.

In the first method, which is not extensively adopted on account of the accurate jigs that are involved, the lens blanks are ground to the final diameter to within a limit of about minus 1/20 mm. Several piles of discs cemented together preparatory to being ground to the correct diameter are shown in *Fig.* 34, which also illustrates the carborundum wheel edging machine.

The edged blanks after being separated and cleaned are mounted in the tool indicated in

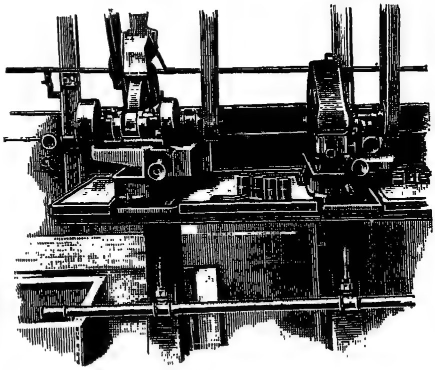


FIG. 34.—Automatic Edging Machine.

Fig. 33, and the curve is then ground true with the periphery, provided the jig and workmanship are sufficiently accurate.

In the more common method, the unedged and partially-shaped blanks are secured individually with pitch within the holder. They are then ground and polished in the usual manner, and after being removed and cleaned

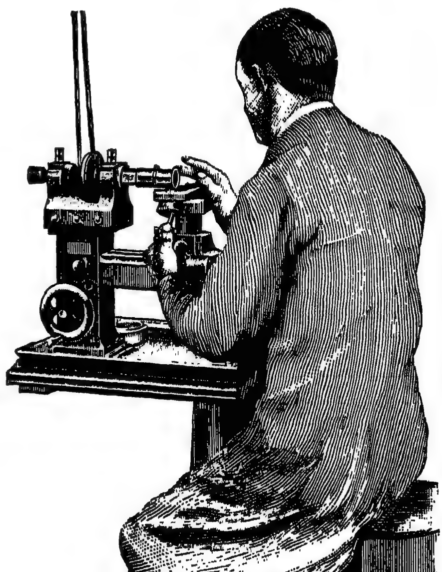


FIG. 35.—Hand Centring and Edging Machine

the periphery of the lens, if circular, is ground true with the optical axis of the lens.

This operation is indicated in *Fig.* 35. The

lens to be edged is slightly heated and cemented to the nose of the hollow spindle of a small lathe head. While the cement is soft the position of the lens is adjusted until the two reflections of a light or other object from the front and back surfaces of the lens[1] do not rotate when the spindle is slowly revolved. This occurs when the optical axis of the lens coincides with the axis of rotation. The cement is then allowed to harden, and the grinding of the periphery true with the axis of rotation, and therefore with the optical axis of the lens, is performed by pressing a brass plate against the edge while using emery or carborundum as an abrasive.

Various types of machines are employed for special operations and more particularly for the production of spectacle lenses,[2]

Astigmatic lenses are produced on machines in which a cylindrical tool is made to oscillate about the axis of the cylinder, the work itself being prevented from rotating. For the multiple production of toric lenses in which the two axes of curvature are approximately at right angles to each other, the lenses are worked upon cylinders of the appropriate radius and ground by means of tools whose cross-sections correspond with the curvatures.

Single and special toric lenses are ground individually upon small special machines, but in an article such as this, which is concerned more with general principles than with the mechanical appliances, it is not possible to describe in detail the numerous types of machine that are employed for spectacle work. The most complete information on this particular subject is obtainable from the catalogues of the optical machine tool makers or from the practical handbooks already referred to.

J. W. F.

Optical Path Difference: the difference in the total optical path travelled by rays of light which pass through an optical system at different distances from the axis. See "Microscope, Optics of the," §§ (7) and (8); also "Light, Rectilinear Propagation of."

OPTICAL ROTATION AND THE POLARIMETER

The rotation of the plane of polarisation was first observed in the year 1811 by Arago in quartz plates. Further, Biot, and also Seebeck, discovered the optical activity of some organic substances, oil of turpentine, and solutions of sugar and of tartaric acid.

Biot established the nature of the phenomenon and its laws in a long series of important papers which extend over a period of forty-seven years (1813 to 1860).

Fresnel published the theory of the properties of quartz and introduced the term "circular polarisation" in 1831.

W. Herschel and J. Herschel recognised the significance of certain faces on the quartz crystal for predicting the direction of the optical rotation.

An epoch-making discovery was Pasteur's in 1848. He proved in the case of tartaric and racemic acids that one and the same active substance may occur in two forms, with opposite rotatory power, as well as in an inactive modification.

Pasteur stated that substances which are optically active as crystals or in solution crystallise in hemihedral forms, *i.e.* the crystal and its mirror cannot be superposed. (Pasteur's rule is not reversible, hemihedral crystals do not always rotate polarised light) (1).

Pasteur thought that the optical activity might be caused by a lack of symmetry of the entire molecule. He imagined that the atoms might be arranged in the form of an irregular tetrahedron or of a screw.

No progress of theoretical importance was made until 1874, when Van t'Hoff and Le Bel, independently of each other, put forward the celebrated theory that the optical rotation of organic compounds is due to the presence of an asymmetric carbon atom. They each assumed that the four valencies of the carbon atom are directed to the points of a regular tetrahedron. If each of these valencies is attached to a different atom or radicle it is seen that two non-superposable tetrahedra result, and these are the left- and right-handed forms of the active substance

Thus the polarimeter has played as celebrated a role in the development of theoretical organic chemistry as its essential parts, the nicol prism and the quartz plate, have done in theoretical optics.

The rapid progress of synthetical organic chemistry in the later decades of the nineteenth century led to the study of a large number of new compounds which rotated polarised light.

One of the most celebrated of the prolonged researches connected with the polarimeter was Emil Fischer's on the constitution of the sugars, in which the interpretation of optical activity played a prominent part.

The discovery of compounds whose optical activity was due to the asymmetry of a nitrogen atom was followed by the brilliant work of Pope on optically active tin, etc., compounds.

Theoretically, optical rotation could arise from asymmetric valencies of elements other than carbon. These predictions have been realised. We are now acquainted with sub-

[1] *Dioptricks* 1692; Molyneux, chap. iv. of *Mechanick-Dioptricks*, p. 220.
[2] *Schule der Optik*, Gleichen und Klein, 1914, Praktischer Teil.

stances whose rotation is due to asymmetric nitrogen, tin, sulphur, selenium, phosphorus, silicon, chromium, and cobalt (2), English investigators, Pope, Peachey, Smiles, Neville and Kipping being chiefly concerned with the first six elements, while Werner was successful in obtaining the active compounds of chromium and cobalt.

Besides these inquiries of fundamental importance, the polarimeter was employed in a classical piece of research by Wilhelmy in 1850. He measured the rate of inversion of cane sugar and put forward the first correct mathematical treatment of the velocity of a chemical reaction. The extraordinary convenience of analysing a solution by optical means without altering its composition was the motive for employing the instrument.

Much work has been carried out on rotatory dispersion, but the interesting results obtained still await theoretical explanation. Technical use was made of rotatory dispersion in 1910 by M. E. Darmois, who found that the proportions of alpha and beta Pinene in rectified oil of turpentine could be determined by measuring the rotatory dispersion (3).

Magnetic rotation of the plane of polarisation, discovered by Faraday, was the subject of prolonged research by W. H. Perkin, senior, who showed that it was a property depending mainly on constitution, although there are some additive relationships.

While the study of rotatory dispersion prevented a too narrow interpretation being placed on the results with sodium light, a study of the influence of the solvents showed that not only the amount but also even the sign of the optical rotation could vary when the same substance was dissolved in different solvents.

Arising out of this purely scientific work is an interesting analytical method. The quantitative estimation of benzene in cyclohexane is difficult and troublesome, but by taking advantage of the fact that benzene is almost without influence on the rotation of ethyl tartrate, whilst cyclohexane exerts a considerable depressing influence, the proportions of the two substances present in a mixture may be estimated within about 3 per cent by the simple determination of the rotatory power of a mixture with a fixed proportion of the ester (4).

The influence of the solvent is also clearly shown in an important paper by F. H. Carr and W. C. Reynolds (5) on the rotatory power of alkaloids. They found, *inter alia*, that hydrastine has a strong dextro-rotation in 50 per cent alcohol, while it is inactive in 95 per cent alcohol, and laevo-rotatory in absolute alcohol. These observations suggest that indirect determinations of inactive substances may often be possible, by measuring their influence on the rotation of an admixed active substance. It has been known for a long while that boric acid and also acetone and other substances have a powerful effect on the rotation of tartaric acid. For the two substances mentioned, there are excellent chemical methods available; where this is not the case a polarimetric method might well be looked for.

Although a distinction may be broadly made between essential oils, which are optically active, and mineral and fatty oils, which are optically inert, the exceptions are of great importance. Castor oil is the only common vegetable oil which exhibits a distinct though slightly variable rotation. Mineral oils were thought for some years to be optically inactive, but further study has proved that they possess a small rotatory power, and this fact has supported one of the theories of the origin of petroleum.

The temperature correction may be of great importance for the analysis of a mixture, thus a mixture of dextrose (+) and invert sugar (+ and −, the latter predominating) may be analysed at 87° C., when the rotation of invert sugar becomes 0.

Starches can be determined by the polarimeter by Ewen's method and its later modifications; thus the products of another important industry may be controlled by this instrument.

The relation between concentration and specific rotatory power has been determined for a number of substances with a very high degree of accuracy, notably cane sugar, galatose, cocaine, lactose, maltose, glucose, camphor, nicotine. "Synthetic camphor" can be distinguished from natural camphor, and adulteration of inactive oils (such as almost all the fatty oils) with rosin or rosin oil is easily detected.

The left- and right-handed forms of active substances, *e.g.* sugars, can be separated by the action of micro-organisms, such as yeast and moulds which preferentially destroy one of the isomers. The isomers differ also in their behaviour with digestive enzymes of the animal body.

The explanation generally adopted is that the enzymes of fermentation and digestion are themselves asymmetric and attach themselves to one optical isomer, in virtue of spatial arrangements which have been compared to the fitting of a key in a lock.

"The first step in what may prove to be an inquiry of considerable significance in biology is marked by the preparation of *d*- and *l*-forms of simple dyes containing an asymmetric system. The work has not proceeded far, but evidence has already been obtained that these optical isomerides are selectively absorbed by wool, and the prospect is thus opened out that they may ultimately be used in the

staining of sections so as to reveal more completely the chemical constitutions of tissues. This field of research has not been explored by the chemist, and there is ample scope for future developments of great importance" (6).

F. T.
J. N. G.

REFERENCES

1. Landolt, *Das optische Drehungsvermögen*, 1898, p. 6.
2. Frankland, *Journal of the Chemical Society*, T, 1912, ci. p. 655.
3. Darmois, *Polarisation rotatoire*, 1910.
4. Patterson and Fleck, *Journal of the Chemical Society*, T, 1910, xcvii. 1773.
5. Carr and Reynolds, *Journal of the Chemical Society*, T, 1910, xcvii. 1328.
6. Porter and Hirst, *J. Am. Chem. Soc.*, 1919, xli. 1264; *J.C.S.* A, 1919, i., cxvi. 558; *Annual Reports Chemical Society*, 1919, p. 75.

OPTICAL SQUARE: a device for deviating a ray of light through 90° irrespective of the angle of incidence. See "Rangefinder, Short-base," § (9).

OPTICAL THICKNESS: the product μt, where μ is the refractive index and t the thickness of a plate of refracting substance. If μ is not constant throughout the material the optical thickness is the line integral of μdt along the path traversed by the light. See "Interferometers, Technical Applications," § (3).

OPTOMETRY: a term sometimes applied to all ocular methods of estimating the refraction of the eye, but generally confined to instruments where an adjustment of lenses is made by the patient, in order to obtain the clear image of an object, and the result recorded on the metric scale. See "Ophthalmic Optical Apparatus," § (11).

ORGAN, REED PIPES IN THE. See "Sound," § (32).

ORGAN PIPES WITHOUT REEDS. See "Sound," § (33).

OXIDE OF IRON: a cause of colour in glass. See "Glass, Chemical Decomposition of," § (1).

OXIDE OF LEAD: a constituent of glass. See "Glass, Chemical Decomposition of," § (1).

— P —

PARABOLIC REFLECTOR. See "Projection Apparatus," § (6).

And electric search-lights. See *ibid.* § (8).

And electric signalling lamps. See *ibid.* § (9).

Used as a headlight for a motor-car. See *ibid.* § (7).

PARALLAX (CHROMATIC): a parallax effect due to the chromatic aberration of the eye. See "Eye," § (27).

PARAXIAL RAYS: rays of light comprised within a narrow pencil close to the axis of an optical system. See "Microscope, Optics of the," § (3). See also "Optical Calculations," § (5).

PARTIAL FLASH: a term applied to the light afforded by a projector when only certain parts of the front aperture are seen as of the same intrinsic brightness as the source; this term applies also to systems in which certain areas of the front aperture are seen filled with a "coloured flash" due to the "partial flashing" of certain primary colours. See "Projection Apparatus," § (3).

PELORUS OR BEARING PLATE, description and method of use. See "Navigation and Navigational Instruments," § (17) (ii.).

PENETROMETER: an instrument for measuring the penetrating power or quality of X-rays. See "Radiology," § (21).

PENTANE LAMP: a flame standard of 10 candle-power devised by Vernon-Harcourt. See "Photometry and Illumination," § (6).

PERIMETER OR SCOTOMETER: an ophthalmic instrument used to measure the field of vision from the macula to the peripheral parts of the retina, and to plot out the portions of the retina which are totally or partially atrophied. See "Ophthalmic Optical Apparatus," § (5).

PERISCOPES

§ (1) PRINCIPLE OF PERISCOPES.—As the name denotes, periscopes are optical instruments in which the general course of the rays, instead of being confined approximately to the neighbourhood of a single straight line, is deflected once or more into new directions with the aim of giving an observer a view from a position in which it is inconvenient to place his head. Among the best known examples are various forms of trench periscope, which consist essentially of small telescopes with reflectors at the top and bottom, or even of a small tube only to which mirrors have been attached at each end, so that the ray paths tend to conform to the shape of a crude letter Z. By the aid of such an instrument the observer is enabled to see what takes place, say, on the far side of a parapet behind which he himself is protected from bullets. Essentially similar in their general aim are the periscopes used on submarine boats to take observations above the sea while the vessel remains submerged. These naval instruments are more complex than those used on land, and as they provide

good illustrations of all the features which are found in periscopes generally, it will be convenient to consider the submarine periscope as the typical instrument of this class.

§ (2) Reflecting Systems.—Although periscopes have the bent ray path as their most obvious characteristic, optically this feature is of no importance. Any refracting telescope can be converted into a periscope by the insertion of mirrors in suitable positions without the necessity for any modification in the construction of the lenses. If reflecting prisms are substituted for mirrors the necessary changes in the system are very slight in character, and may be allowed for in the original design by supposing symmetrical blocks of glass of suitable thickness to be present, as no optical difference would then exist between the instrument with a straight ray course and that in which the light path suffers deviations. The reflecting and refracting systems are independent also in another sense, for though the size and the accuracy of surface necessary in the mirrors may vary greatly with their positions relative to the lenses, yet those positions are unimportant as far as the type of imagery yielded by the complete instrument is concerned. For instance, if a certain combination of mirrors and a telescope produces an upright image of a clock face with the hours increasing on a right-handed rotation, no image either inverted or with left-handed rotation will be produced by transferring the whole reflecting system so that it is entirely external to the telescope, and either precedes or follows it. To decide therefore what mirror systems are of value for the construction of periscopes, it is only necessary to consider whether the telescope to be employed is of the inverting or of the erecting type.

The commonest reflecting system consists essentially of two plane reflecting surfaces parallel to one another (*Fig.* 1). Light successively reflected at these two surfaces emerges parallel to its original direction, being merely laterally displaced by an amount proportional to the separation of the mirrors and to the sine of the angle of incidence. A clock face examined by light reflected from such mirrors presents the same appearance as when viewed direct. Such a system of mirrors in conjunction with an erecting telescope gives a correct view of the scene examined when the observer faces the direction of sight. Another simple system consists of two plane mirrors at right angles to one another (*Fig.* 2). Since the deviation produced in any ray is twice the angle between the outward drawn normals to the mirrors, the observer must have his back towards the direction of sight, and from this it follows that an image will appear correct to the observer when it is upright, but inverted left for right in space relative to the object. Now the two mirrors by themselves, if placed one above the other with their lines of intersection horizontal, will obviously produce inversion in the vertical but not in the horizontal direction; Since an inverting telescope inverts in both directions, the combination of these two mirrors with an inverting telescope will yield images of the type required (*Fig.* 3). If a correct view is required with an erecting telescope when the observer has his back to the direction from which light reaches the instrument, the reflecting system must be of a type which returns an image of correct appearance to the observer when used alone. The best known system of this kind consists of three plane mirrors mutually at right angles to one another, and many others only slightly less simple may be devised. A periscope of the reversed vision type may therefore be constructed with a plane upper mirror deflecting the light downwards through an erecting telescope at the lower end of which is a "roof" prism or two mirrors forming a roof, the edge of which is normal to the first mirror (*Fig.* 4). Roof prisms would be utilised very widely in optical instruments for their valuable properties were it not that the high degree of accuracy required in the angle between the two reflecting surfaces to avoid double images makes their manufacture very costly. In many cases this difficulty can be avoided at the cost of compactness by displacing the

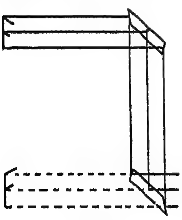


Fig. 1.

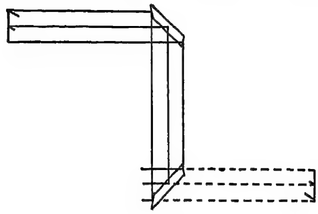


Fig. 2.

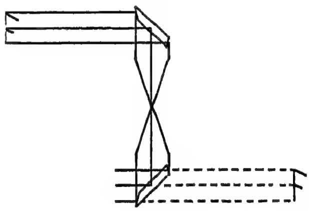


Fig. 3.

roof combination laterally, so that the light always meets the surfaces in the same order. Another case in which a roof combination

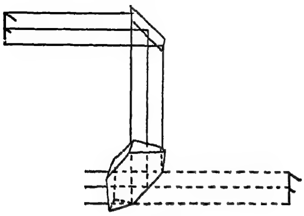


FIG. 4.

may be used is when an inverting telescope is employed with the observer facing the view. If the roof prism replace the lower plane prism of *Fig.* 1, the lateral inversion it causes compensates the inversion of the telescope in the horizontal plane, but the vertical inversion is uncorrected. The substitution at the upper end of the periscope of two mirrors inclined to one another in the vertical plane at an angle of 45° for the single mirror is a possible method of completing the rectification of the image (*Fig.* 5). An alternative of much importance consists in the introduction of an inverting isosceles prism. With this combination an erect image may be secured conveniently whatever the angle between the line of sight and the direction of observation

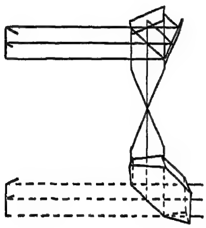


FIG. 5.

may be. The best known instrument of this kind has an upper isosceles right-angled prism, a lower roof prism, and an isosceles inverting prism placed along the vertical axis of the telescope, so mounted that it rotates through half the relative displacement of the upper and lower prisms. This arrangement is illustrated in *Fig.* 6.

In submarine periscopes it is considered important that the directions of sight and of observation should be identical, and as the use of roof reflectors is not free from difficulties erecting telescopes become necessary.

§ (3) FIELD OF VIEW AND ILLUMINATION.—From the point of view of light deflection then the periscope offers no new problem at all for consideration. Optically its special characteristic is that, when the extent of the field of view required and the amount of light necessary from each part of the field are considered, the instrument is of such unusual length that special forms of construction are essential to enable these particular features to be realised in an acceptable manner.

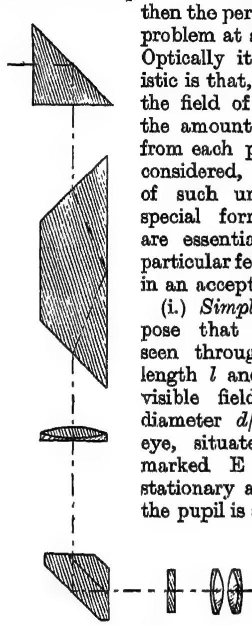


FIG. 6.

(i.) *Simple Periscopes.*—Suppose that a distant view is seen through a long tube of length l and diameter d. The visible field has an angular diameter d/l if the observer's eye, situated in the position marked E in the figures, is stationary and the diameter of the pupil is small in comparison with d (*Fig.* 7). If a lens of diameter d and focal length $l/6$ is placed in the tube at one-third its length from the observer's end, an inverted image of the distant view of angular diameter $3d/l$ will be seen at the principal focus of the lens which is at a distance $l/6$ from the observer's eye (*Fig.* 8). As

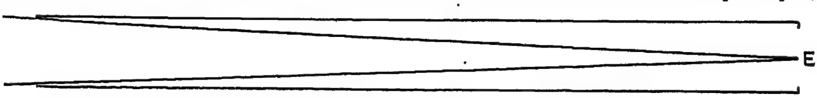


FIG. 7.

far as the field of view is concerned the eye is virtually placed at e, the image of E formed by the lens. As exception is taken to a system such as the above, which necessitates focussing

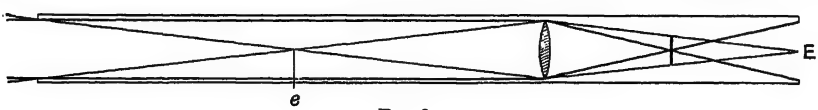


FIG. 8.

the eye on a fixed near plane, the single lens must be replaced by a telescopic system. The simplest consists of two single lenses of equal focal length with their separation double the focal length of either. If these are placed

at the ends of the tube the field of view is the same as for the vacant tube, but the image is inverted (*Fig.* 9). If the lenses are now reduced in focal length and are appropriately placed in symmetrical positions in the tube, the field can be increased up to four times its original value when the lens separation is $\frac{1}{2}l$ (*Fig.* 10). If the symmetrical arrangement is abandoned, the most favourable disposition with two equal lenses is reached when the focal length is $l/5$, the distance of the first lens from the object end of the tube being $2l/5$, which is also the distance between the lenses (*Fig.* 11). The image of the eye is then half-way between the front of the tube and the leading lens, and it is easy to see that a variation of the system in one direction results in a reduction of the field below $5d/l$ through the failure of the rays required to meet the aperture of the first lens, and a variation in the other direction leads to the obstruction of rays between the object and the virtual position of the eye by the front of the tube. To take an actual example, if the tube is thirty feet in length and six inches in diameter, the natural field of view is about one degree, and the best with two lenses is about five degrees. Fields of this magnitude fall far short of those required in submarine periscopes, and the inverted image is also objectionable. The addition of further lenses enables both defects to be removed.

A particularly simple form of erecting telescope consists of three equal lenses spaced at equal intervals, the separations being three times the focal length of the individual lenses. If such a system were inserted in the tube and occupied its entire length, the eye would be virtually placed a distance $\frac{1}{2}l$ in front of the tube, and the field of view would be $2d/l$ (*Fig.* 12). As the outer lenses are made to approach the central lens and the focal lengths of all suitably reduced, the field of view increases, reaching the best value $7d/l$ when the separations and the distance from the front of the tube to the first lens are each $2l/7$ (*Fig.* 13). This is not the only form of telescope of unit magnifying power which can be constructed of three equal lenses uniformly spaced, for evidently in *Figs.* 9 to 11 a third lens of any power could be inserted half-way between the lenses shown in the plane of the real image without disturbing the telescopic character of the image. The effect of such a lens is merely to alter the extent of the field of view. The

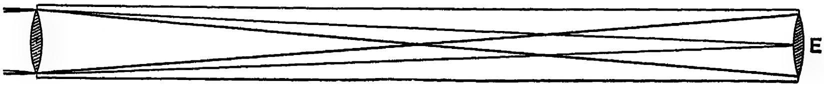


FIG. 9.

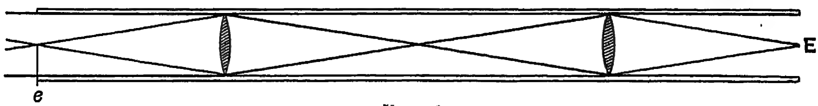


FIG. 10.

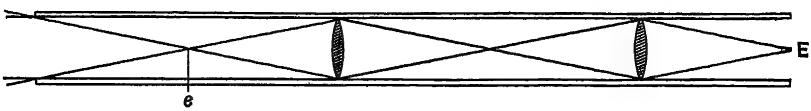


FIG. 11.

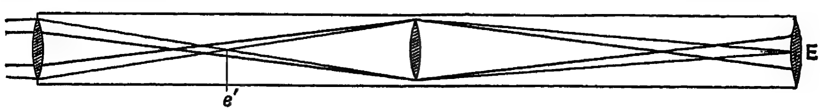


FIG. 12.

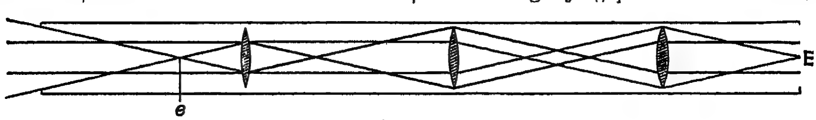


FIG. 13.

insertion of the lens in this system, however, does not influence the inversion of the image, so that the one system of three lenses presents an upright and the other an inverted image.

(ii.) *Six-lens Periscopes.*—When a greater number of lenses is used there is a choice of many different separations when the spacing is uniform and the lenses are of equal focal length. Consider, for example, a system with six lenses. Two telescopes each having three lenses if combined will produce a six-lens telescope with an upright image, whether the image is upright or inverted with three lenses. On the principle of combining two telescopes together it is obvious that with twelve lenses all these separations would give erect images. In addition, with twelve lenses there would be six separations resulting in inverting telescopes.

The five forms of telescope each containing six lenses are illustrated in *Figs.* 14 to 18 and the number of inversions is different in every case, the number for each separation being given in brackets in the above table. Other things being equal, the form in which f is large in comparison with t is the most desirable, since this implies lenses with slight curvatures.

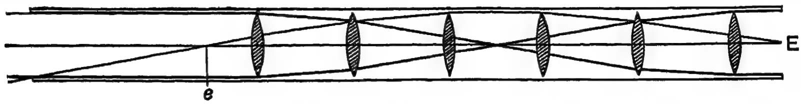


FIG. 14.

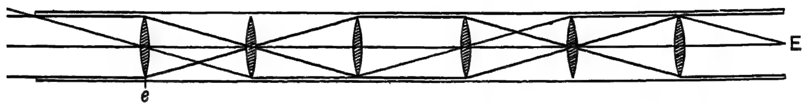


FIG. 15.

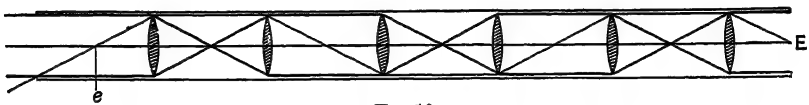


FIG. 16.

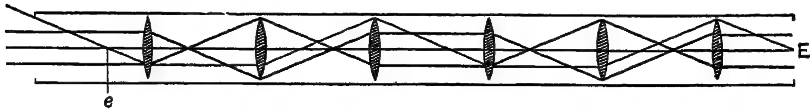


FIG. 17.

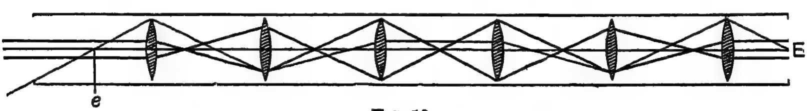


FIG. 18.

Three telescopes of two lenses each will evidently combine to form an inverting six-lens telescope. In addition there are two new forms, both inverting, which cannot be divided into separate telescopes. Thus, with six lenses each of focal length f the separations t may have the values given in the following table:

TABLE I

Inverting Telescopes.	Erecting Telescopes.
$t=(2-\sqrt{3})f$ (1)	$t=f$ (2)
$t=2f$ (3)	$t=3f$ (4)
$t=(2+\sqrt{3})f$ (5)	

But it is by no means the case that other things are equal, for the systems differ from one another in the extent of the field of view, in the brightness of the image at the centre of the field, and in the way in which the brightness varies from point to point of the field. The system with one inversion, illustrated in *Fig.* 14, is the most desirable as regards the curvatures of the lenses, but since the virtual eye position is the farthest of the series from the leading end of the tube, this form is the least satisfactory for the extent of the field of view. The systems with three and five inversions, illustrated in *Figs.* 16 and 18 respectively, are those with the largest fields of view, the two being on an equality. Not far behind them is the system with four inversions shown in *Fig.* 17.

(iii.) *Illumination.*—In order to consider the illumination at the centre of the field of view, the extreme rays of an incident beam parallel to the axis are represented in the diagrams by thicker lines than those used to indicate the extent of the field of view. The systems with one, two, and three inversions are on an equality, and transmit to the rear of the tube beams of light which on entrance and exit completely fill its diameter. When the number of inversions is increased to four, and still more when there are five, the width of the useful axial incident beam is much reduced, and the last two forms are not nearly as satisfactory as those which precede them for central illumination. It would thus appear that the greatest advantages rest with the form shown in *Fig.* 16, but when the illumination in the outer parts of the field of view is considered, it is seen that this form has distinctly undesirable features. The forms with four and five inversions alone show no decided change in illumination, as the origin of the beam of incident rays departs from the axis of the tube, but beyond certain angles the brightness of the image is diminished through light striking the interior of the tube as in the other forms. At the extreme angular field the intensity of the light falls to zero with the forms of *Figs.* 14, 16, and 18, but is finite with the remaining forms.

Figs. 14 to 18 are special cases of their kinds, as the extreme rays which determine the fields of view have been assumed to follow paths similar to those of rays which limit the axial beam of parallel rays. Other results as regards the field of view or the variation in illumination are obtained by assuming that the entire system of lenses is displaced along the tube, the eye remaining at the end of the tube. It will suffice to consider two other cases, in which the lenses of the two erecting telescopes are displaced towards the eye. *Fig.* 19 represents the same system as *Fig.* 15, but all the lenses have been moved towards the eye through a distance equal to half their focal length. The field of view is evidently the same as before, but the amount of light transmitted at the extreme angle is now less than before. Again, in *Fig.* 20 a system similar to that shown in *Fig.* 17, but with rather longer focal lengths for the lenses, has been displaced towards the eye through a distance equal to half the focal length of each lens. The field of view is now increased to an equality with those of the systems with three and five inversions, but, as in these other cases, the illumination at the edge of the increased field falls to zero. It is clear that a much more detailed analysis is necessary to determine exactly where the balance of advantage falls.

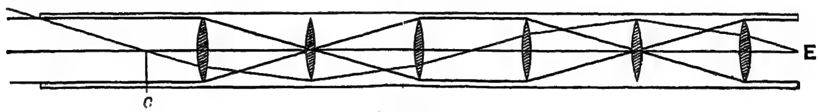


FIG. 19.

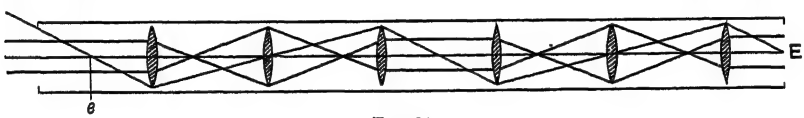


FIG. 20.

§ (4) THEORETICAL CONSIDERATION.—The diagrams just considered show that the field of view is enlarged by causing any ray of light, instead of travelling straight along the tube, to be deviated by a series of lenses from side to side. The number of lenses to an oscillation is arbitrary, and when the number of lenses to an oscillation is large, as in *Fig.* 14, the ray path resembles a sine curve. If there are n lenses to the telescope, each of power κ, the separation of successive lenses being t, the power of the complete system $\kappa_{1,n}$ is derived from

$$\begin{aligned}\kappa_{1,n} = \kappa_1 + \kappa_2 + &\ \ldots\ + \kappa_n \\ &- \kappa_1 t_1 \kappa_2 - \kappa_1(t_1 + t_2)\kappa_3 \ \ldots \\ &- \kappa_1(t_1 + t_2 + \ \ldots\ + t_{n-1})\kappa_n \\ &- \kappa_2 t_2 \kappa_3 - \kappa_2(t_2 + t_3)\kappa_4 \ \ldots \\ &+ \kappa_1 t_1 \kappa_2 t_2 \kappa_3 + \ \ldots,\end{aligned}$$

as is well known. On putting all the κ's and t's equal to one another, the coefficients depend on the numerical value of the sums of continued products obtained by dividing at $r-1$ points a line whose length varies from r to n into r parts each of a length represented by an integer. The result is

$$\begin{aligned}\kappa_{1,n} = n\kappa - \frac{(n+1)n(n-1)}{3!}\kappa^2 t \\ + \frac{(n+2)(n+1)n(n-1)(n-2)}{5!}\kappa^3 t^2 \ \ldots \\ + (-1)^r \frac{(n+r)!}{(n-r-1)!(2r+1)!}\kappa^{r+1} t^r \ \ldots,\end{aligned}$$

so that the system is telescopic if κt is made equal to one of the roots of the equation

$$\Sigma_r(-1)^r\frac{(n+r)!}{(n-r-1)!(2r+1)!}x^r=0,$$

and, like the ray path, this equation, if written in the form

$$\Sigma_r(-1)^r\frac{n(n^2-1^2)(n^2-2^2)(n^2-3^2)\dots(n^2-r^2)}{(2r+1)!}x^r=0,$$

is distinctly reminiscent of circular functions. This suggests that the circular functions will occur prominently in the theory of systems of lenses suitable for the transmission of light down a tube of constant diameter, and it is readily seen that if the four Gaussian constants are suitable circular functions the system may have such properties as those found in the systems already illustrated. If these constants are denoted for brevity by A, B, C, D, where

$$A=\kappa_{1,n},\ B=\frac{\partial\kappa_{1,n}}{\partial\kappa_1},\ C=\frac{\partial\kappa_{1,n}}{\partial\kappa_n},\ D=\frac{\partial^2\kappa_{1,n}}{\partial\kappa_1\partial\kappa_n},$$

the relation $BC-AD=1$ suggests cosine expressions for B and C and sine expressions for A and D. Moreover, the essential equations for the combination of two systems are

$$\left.\begin{aligned}A_{1,2}&=A_1B_2+A_2C_1\\B_{1,2}&=B_1B_2+A_2D_1\\C_{1,2}&=A_1D_2+C_1C_2\\D_{1,2}&=B_1D_2+C_2D_1\end{aligned}\right\}, \quad . \quad . \quad (1)$$

where the quantities are reckoned for a common terminal point to the two systems. For the distances x, x' from the selected terminal points of a pair of conjugate points satisfy the relation

$$Axx'+Bx'-Cx-D=0,$$

where x and x' are positive when measured from the terminal points in the direction in which light travels through the system. Thus, in this case, if x and x' refer to a first system which is to be combined with a second system for which the distances are represented by x' and x'', the two values obtained for x' must be identical, so that

$$\frac{C_1x+D_1}{A_1x+B_1}=-\frac{B_2x''-D_2}{A_2x''-C_2},$$

or

$$(A_1B_2+A_2C_1)xx''+(B_1B_2+A_2D_1)x''-(A_1D_2+C_1C_2)x-(B_1D_2+C_2D_1)=0,$$

and this must be equivalent to

$$A_{1,2}xx''+B_{1,2}x''-C_{1,2}x-D_{1,2}=0.$$

Also

$$\begin{aligned}(B_1B_2+A_2D_1)(A_1D_2+C_1C_2)&\\-(A_1B_2+A_2C_1)(B_1D_2+C_2D_1)&\\=(B_1C_1-A_1D_1)(B_2C_2-A_2D_2)&\\=1=B_{1,2}C_{1,2}-A_{1,2}D_{1,2},&\end{aligned}$$

so that

$$\begin{aligned}A_1B_2+A_2C_1&=\pm A_{1,2},\\B_1B_2+A_2D_1&=\pm B_{1,2},\\A_1D_2+C_1C_2&=\pm C_{1,2},\\B_1D_2+C_2D_1&=\pm D_{1,2},\end{aligned}$$

and by considering a special case, such as a thin lens in which $B_2=C_1=1$, $A_{1,2}=A_1+A_2$, or a lens in which A_2 is gradually increased from zero so that $A_{1,2}$ has the same sign as A_1, it is seen that the upper sign must be adopted. If now l denotes a length it may be supposed that A_1, B_1, C_1 are defined by the angles α_1, β_1, γ_1, where

$$A_1=\frac{\sin\alpha_1}{l},\ B_1=\frac{\cos(\alpha_1-\beta_1)}{\cos\beta_1},\ C_1=\frac{\cos(\alpha_1-\gamma_1)}{\cos\gamma_1},\ (2)$$

and it then follows from the identity $BC-AD=1$ that

$$D_1=-\frac{l\sin(\alpha_1-\beta_1-\gamma_1)}{\cos\beta_1\cos\gamma_1}. \quad . \quad (2a)$$

If now two systems with the same l are combined the addition equations give

$$A_{1,2}=\frac{\sin(\alpha_1+\alpha_2)}{l}+\frac{\sin\alpha_1\sin\alpha_2}{l}(\tan\beta_2+\tan\gamma_1),$$

$$B_{1,2}=\frac{\cos(\alpha_1+\alpha_2-\beta_1)}{\cos\beta_1}+\frac{\cos(\alpha_1-\beta_1)\sin\alpha_2}{\cos\beta_1}(\tan\beta_2+\tan\gamma_1),$$

$$C_{1,2}=\frac{\cos(\alpha_1+\alpha_2-\gamma_2)}{\cos\gamma_2}+\frac{\sin\alpha_1\cos(\alpha_2-\gamma_2)}{\cos\gamma_2}(\tan\beta_2+\tan\gamma_1),$$

$$D_{1,2}=-\frac{l\sin(\alpha_1+\alpha_2-\beta_1-\gamma_2)}{\cos\beta_1\cos\gamma_2}+\frac{l\cos(\alpha_1-\beta_1)(\cos\alpha_2-\gamma_2)}{\cos\beta_1\cos\gamma_2}(\tan\beta_2+\tan\gamma_1),$$

and the combined system has expressions for the Gaussian constants of the same form as the components if

$$\tan\beta_2+\tan\gamma_1=0,$$

that is, if $\qquad \beta_2+\gamma_1=0.$

The α of the compound system is then the sum of the α's of its components, while its β and γ are respectively β_1 and γ_2, that is, the external angles associated with its components. Moreover, if a single lens is considered referred to its unit points as terminal points, A is arbitrary, B and C are both unity, and, if the lens is thin, D is zero. Since α cannot be zero without making A zero, it follows that for such a lens

$$\alpha-\beta=\beta,\ \alpha-\gamma=\gamma,$$

or $\qquad \beta=\gamma=\tfrac{1}{2}\alpha.$

The preceding equation for $A_{1,2}$ may be interpreted in terms of each component lens referred to its own principal points; if t is the separation of the internal principal points of the two components this gives

$$\frac{\sin(\alpha_1+\alpha_2)}{l}=\frac{\sin\alpha_1}{l}+\frac{\sin\alpha_2}{l}-t\,.\,\frac{\sin\alpha_1}{l}\,.\,\frac{\sin\alpha_2}{l},$$

or $$t=l(\tan\tfrac{1}{2}\alpha_1+\tan\tfrac{1}{2}\alpha_2).$$

It follows that in a system of n lenses in which the power of the rth lens is

$$\kappa_r=\frac{\sin\alpha_r}{l}, \qquad . \quad . \quad . \quad (3)$$

and the separation between the rth and $(r+1)$th lenses is

$$t_r=l(\tan\tfrac{1}{2}\alpha_r+\tan\tfrac{1}{2}\alpha_{r+1}), \qquad . \quad . \quad (4)$$

the four Gaussian constants referred to the external principal points of the extreme components are given by

$$\left.\begin{aligned}
A&=\frac{\sin\alpha}{l}\\
B&=\frac{\cos(\alpha-\frac{1}{2}\alpha_1)}{\cos\frac{1}{2}\alpha_1}\\
C&=\frac{\cos(\alpha-\frac{1}{2}\alpha_n)}{\cos\frac{1}{2}\alpha_n}\\
D&=-\frac{l\sin(\alpha-\frac{1}{2}\alpha_1-\frac{1}{2}\alpha_n)}{\cos\frac{1}{2}\alpha_1\cos\frac{1}{2}\alpha_n}
\end{aligned}\right\}, \qquad . \quad (5)$$

where

$$\alpha=\alpha_1+\alpha_2+\ .\ .\ .\ +\alpha_n.$$

If α is a multiple of π the system is evidently a telescope of unit power, inverting if the multiple is odd, erecting if it is even.

§ (5) Geometrical Theory.—Equations (3) and (4) show that the properties of the special kind of system which has been considered have a parallel in polygons described about a circle. If radii are drawn (*Fig.* 21) making

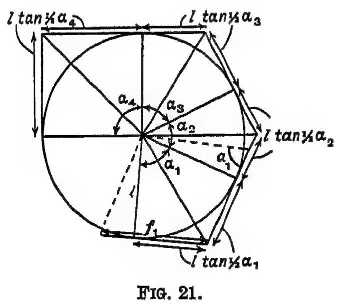


Fig. 21.

with one another successive angles $\alpha_1, \alpha_2, \alpha_3 \ldots$, the sides of the polygon which touch the circle where these radii meet the circumference will represent the lengths of the sections into which the lenses divide the periscope if the radius of the circle is l. The focal lengths of the lenses are represented by the sides of a rhombus having two sides along successive sides of the polygon and the centre of the circle as the opposite vertex. By constructing polygons in this way the properties of any system are investigated more simply than by drawing detailed diagrams such as *Figs.* 14 to 20.

It is convenient to extend this construction and replace the polygon by a right prism of uniform height. The justification for the use to be made of this extension follows from the best known properties of simple lenses. Referring to *Fig.* 22, let O be a point equidistant

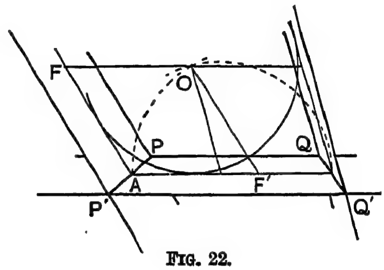
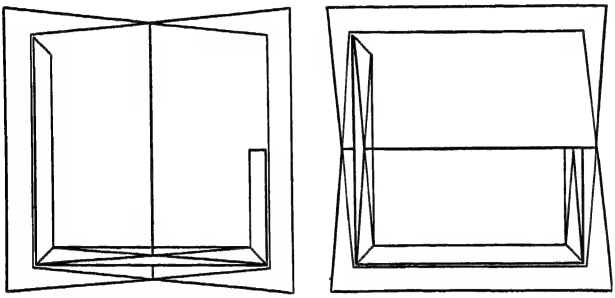


Fig. 22.

from two planes which intersect in the straight line PP′. Suppose that A is the foot of the perpendicular to PP′ from O, and that FA, AF′ lie each in one of the two planes, OFAF′ being a rhombus in the plane normal to PP′. If the two planes FPP′, F′PP′ are regarded as axial sections of the object space and image space for a thin lens with F, F′ as its principal foci and PP′ for its coincident unit planes, any object point is connected to its image point by a straight line which passes through O. The angle which the planes make with one another is immaterial, and if the image plane is rotated about PP′ the locus of O is a circle described about F as centre. Similarly, if the image plane is fixed and the object plane is revolved about PP′, O will move along a circular path with its centre at F′. If a second lens is introduced which meets the first image plane in QQ′, the effect of the second refraction may be determined through the intermediary of the first image plane by projecting through a suitable point O′ on to a third plane which contains the straight line QQ′. If the second plane is kept fixed and the first and third are rotated, O and O′ describe coplanar circles when the two lenses have a common axis, and under suitable conditions these circles intersect, and projection from this common centre gives all the correspondences that exist. In the cases considered here these conditions are satisfied, and also further projection from the same centre is possible, and the effect of many successive refractions may thus be exhibited in this particularly simple way.

The height of the prism corresponds to the diameter of the tube, that is, to the diameter of the lenses, so that the paths of rays may be depicted on the surface of the prism. It is at once evident that, since projection through the centre gives all correspondences after any number of refractions, the complete path of any ray through the entire system is determined by the intersection of a plane through the centre of the circle with the prism surface. If a plane rotates about an axis through the centre parallel to the direction of the object space axis, the width of the beam which can be transmitted is determined by the intersection of the plane with the points on the prism surface at the greatest distance from the axis. Rotation of a plane about another axis determines the field of view, the axis of rotation in this case passing through the point where it is proposed that the eye shall be placed. Since the instrument is assumed to be a telescope this point must be on a side of the prism parallel to that on which the incident light is represented, and the angular field will depend upon the angle between the lines in which the plane through the centre and the eye point meets the plane for incident rays in its extreme positions; these positions are, of course, determined by the requirement that the plane must everywhere meet the prism surfaces within the limits set by the prism height. It is evident that in practice the axis for the aperture and that for the field of view will be approximately perpendicular. *Fig.* 23 shows in perspective the prismatic diagram with the extreme planes corresponding to *Fig.* 11. In this simple case the diagram shows at once that the separation of the lenses must be twice the focal length, that the axes for aperture and field of view are preferably at right angles, and that the front section of the tube may be equal to that of the middle section and twice as great as the third section.



Aperture Axis and Planes **Field Axis and Planes**

FIG. 23.

In the diagrams corresponding to other simple cases the essential relations between geometrical magnitudes may be derived at once from the properties of the polygon or prism. For example, when the sections are the sides of a regular hexagon the focal length of each lens is equal to the distance separating successive lenses, and when the polygon is an equilateral triangle each separation becomes equal to three times the focal length. It is, of course, unnecessary to assume that all the sections or the focal lengths of all the lenses are equal. The geometry of the polygon shows by itself what relations must subsist between the focal lengths and the separations for a system which is telescopic of power ± 1 to be secured. In any particular case, when a position for the eye has been selected, the extent to which the tube may project beyond the first lens without reducing the extent of the field is determined by producing the side of the polygon which corresponds to the object space until a point is reached as far from the field axis as any other point on the fundamental polygon. The variation in brightness as the field is changed is determined most readily by considering a movement of the eye point parallel to the axis of the prism, so that the field axis is inclined to a principal section, and finding the extent of the field for the displaced eye position. Considered in this way it is at once obvious that the greatest extent of field is visible only when the eye is on the axis of the tube.

Figs. 24 to 30 show the polygonal diagrams corresponding to the six-lens systems of *Figs.* 14 to 20. To render the exact character of the system more evident from the diagrams a small separation has been introduced between the faces, when these should strictly

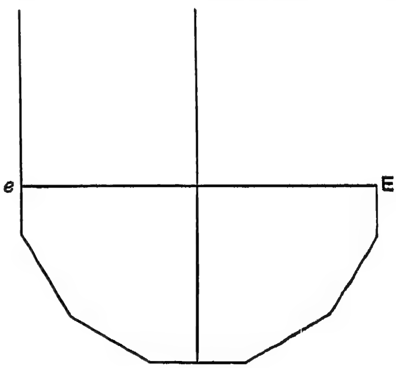


FIG. 24.

be represented by coincident lines. In all cases but *Figs.* 25 and 27, corresponding to systems 15 and 17, the axes for aperture and field of view are perpendicular to one another.

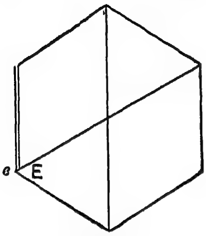


FIG. 25.

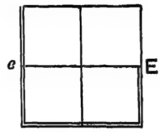


FIG. 26.

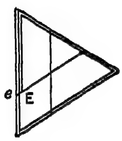


FIG. 27.

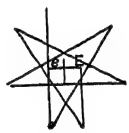


FIG. 28.

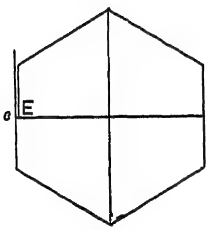


FIG. 29.

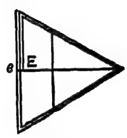


FIG. 30.

§ (6) TUBES OF VARIOUS DIAMETERS.—The preceding theories are chiefly of value in considering the best arrangement of lenses when it is desired to compare alternative systems in a tube of given diameter. The actual diameter of the tube is of great importance apart from the ratio of the diameter to the length, since in actual observation the pupil of the eye should be approximately filled with light, and thus the ratio of the pupil diameter to that of the tube enters into the problem.

The periscope that has been considered may be compared with a geometrically similar instrument in which the diameter of the tube is equal to the diameter of the eye pupil. Both have the same field of view, since this depends on the ratio the transverse dimensions bear to the longitudinal measurements. On the other hand, the periscope with the larger diameter transmits much more light from any given object point, the ratio depending on the areas of the cross sections of the tubes. Some of this will enable the larger tube to present to an eye on its axis a large field of sensibly uniform illumination, while in the smaller instrument the intensity will fall off directly the object point departs from the axis. Much useless light will, however, be transmitted in addition in the larger telescope, and the instrument can be readily modified in a way which enables this unwanted light to be exchanged for an increased field of view, or alternatively for a system containing a smaller number of lenses for a given length of tube. Thus, large diameters are of great assistance in yielding simple solutions of the optical problems involved in periscope design. On the other hand, in nearly all military periscopes great importance has been attached to a small head which renders the periscope inconspicuous. In submarine periscopes these two conflicting requirements are to some extent reconciled by the combination of a short narrow head with a long main tube of large diameter. It thus becomes of interest to inquire how the dimensions of the system vary when a change is made in the aperture which the lenses may have. As a result of the change of aperture the complete instrument when telescopic need not be of unit magnifying power, so that the expressions for B and C must be generalised by the introduction of a further arbitrary quantity. Bearing in mind that the identity

$$BC-AD=1$$

has to be satisfied, it is natural to consider systems in which the four Gaussian constants take the form

$$\left.\begin{aligned} A_1&=\frac{\sin\alpha}{l}\\ B_1&=\frac{p\cos(\alpha-\frac{1}{2}\alpha_1)}{\cos\frac{1}{2}\alpha_1}\\ C_1&=\frac{\cos(\alpha-\frac{1}{2}\alpha_m)}{p\cos\frac{1}{2}\alpha_m}\\ D_1&=-\frac{l\sin(\alpha-\frac{1}{2}\alpha_1-\frac{1}{2}\alpha_m)}{\cos\frac{1}{2}\alpha_1\cos\frac{1}{2}\alpha_m}\end{aligned}\right\}, \quad . \quad (6)$$

so that when the system is telescopic the magnifying power is p. Let these be the constants of a first system to which a second is to be added with values l' and p' in the places of l and p. The second system referred to the terminal point corresponding to the last lens of the first system will then have constants

$$A_2=\frac{\sin\alpha'}{l'},$$

$$B_2=\frac{p'\cos(\alpha'+\frac{1}{2}\alpha_m)}{\cos\frac{1}{2}\alpha_m},$$

$$C_2=\frac{\cos(\alpha'-\frac{1}{2}\alpha_n)}{p'\cos\frac{1}{2}\alpha_n},$$

$$D_2=-\frac{l'\sin(\alpha'+\frac{1}{2}\alpha_m-\frac{1}{2}\alpha_n)}{\cos\frac{1}{2}\alpha_m\cos\frac{1}{2}\alpha_n},$$

and by equations (1) the Gaussian constants of the combination of the two systems are given by

$$A_{1,2}=\frac{\sin(\alpha+\alpha')}{L},$$

$$B_{1,2}=\frac{P\cos(\alpha+\alpha'-\frac{1}{2}\alpha_1)}{\cos\frac{1}{2}\alpha_1},$$

$$C_{1,2}=\frac{\cos(\alpha+\alpha'-\frac{1}{2}\alpha_n)}{P\cos\frac{1}{2}\alpha_n},$$

$$D_{1,2}=-\frac{L\sin(\alpha+\alpha'-\frac{1}{2}\alpha_1-\frac{1}{2}\alpha_n)}{\cos\frac{1}{2}\alpha_1\cos\frac{1}{2}\alpha_n},$$

subject to the relations

$$\left.\begin{aligned}L=pl'=\frac{l}{p'}\\ P=pp'=\frac{l}{l'}\end{aligned}\right\}, \quad . \quad . \quad . \quad (7)$$

the latter of which might have been written down at once as obvious from the meaning of magnifying power. Consider now a system of lenses placed in a tube, of which the first portion is of uniform diameter d, a second portion in which the diameter increases from d to kd, and the third portion of uniform diameter kd. Let the systems which occupy these sections be identified by the numbers 1, 2, 3 respectively. Since the l's are proportional to the sizes of the images that the system forms of any given external objects, and these for the same angular field must vary as the diameter of the tube, the values which should be inserted in (7) are

$$l_{1,2}=l_{1,3}=kl_1, \quad . \quad . \quad . \quad (8)$$

in addition to the conditions that the systems in the outer sections are of the type used in a tube of uniform diameter, that is,

$$p_1=p_3=1. \quad . \quad . \quad . \quad (9)$$

Considering now the combination of systems 1 and 2, equations (7) give as a consequence of (8) and (9)

$$l_2=kl_1, \quad p_2=p_{1,2}=\frac{1}{k}. \quad . \quad . \quad (10)$$

If the second portion is telescopic this merely amounts to the statement that the magnifying power of this telescope is $1/k$. Applying (7) now to the complete instrument, which is divided into two parts (1, 2) and 3, the first equation gives

$$p_{1,2}l_3=l_{1,3},$$

or by (8) and (10)

$$l_3=k^2l_1. \quad . \quad . \quad . \quad (11)$$

This result is of the greatest importance, since it shows that if the diameter is increased k times, and consequently the focal length of the system enlarged in the same ratio, the focal lengths of the individual lenses in the enlarged portions and their separations will be increased k^2 times. Thus, if the diameter of the tube is halved the lenses must be packed four times as closely, and since this means both a greater amount of scattered light and increased astigmatism or curvature and a more dispersed secondary spectrum, the curtailment of the length of the narrow portion of the periscope tube to the minimum possible value is evidently of great importance. *Fig.* 31 illustrates the modification, as the diameter changes, in a system of the type shown in *Fig.* 15.

In actual periscopes the system which occupies the section where the enlargement in the diameter occurs is not usually a telescope, but this is not vital to the argument that has been employed, for portions of the systems which precede or follow may be arbitrarily regarded as parts of the intermediate system. It is, however, always possible to regard this section as a Galilean telescope, though, in fact, no negative lens is employed. The negative lens in this case is supposed to be combined with a lens of the system in the narrower portion of the tube, the two together being either equivalent to a positive lens or preferably to a single lens of zero power, so that no lens is required. In like manner the positive lens may be combined with a lens of the other system to produce a more powerful lens, but this course offers no special advantages. The powers and separations of the lenses on changing the diameter of the tube then are represented by the expressions given in Table II., where lens $m-1$ is the last of the

TABLE II

Powers.	Separations.
$\dfrac{\sin\alpha_{m-2}}{l}$	
	$l(\tan\frac{1}{2}\alpha_{m-2}+\tan\frac{1}{2}\alpha_{m-1})$
$\dfrac{\sin\alpha_{m-1}}{l}$	
	$l\left(\tan\frac{1}{2}\alpha_{m-1}+\dfrac{k-\cos\alpha_m}{\sin\alpha_m}\right)$
$\dfrac{\sin\alpha_m}{kl}$	
	$kl\left(\dfrac{1-k\cos\alpha_m}{\sin\alpha_m}+k\tan\frac{1}{2}\alpha_{m+1}\right)$
$\dfrac{\sin\alpha_{m+1}}{k^2l}$	
	$k^2l(\tan\frac{1}{2}\alpha_{m+1}+\tan\frac{1}{2}\alpha_{m+2})$
$\dfrac{\sin\alpha_{m+2}}{k^2l}$	

smaller diameter and lens m the first of the larger diameter. The four Gaussian constants when lenses $m-1$, m, $m+1$ have just been included are given in Table III. :

TABLE III

n.	$A_{1,n}$.	$B_{1,n}$.	$C_{1,n}$.	$D_{1,n}$.
$m-1$	$\frac{\sin \alpha_{1,m-1}}{l}$	$\frac{\cos(\alpha_{1,m-1}-\frac{1}{2}\alpha_1)}{\cos \frac{1}{2}\alpha_1}$	$\frac{\cos(\alpha_{1,m-1}-\frac{1}{2}\alpha_{m-1})}{\cos \frac{1}{2}\alpha_{m-1}}$	$-\frac{l \sin(\alpha_{1,m-1}-\frac{1}{2}\alpha_1-\frac{1}{2}\alpha_{m-1})}{\cos \frac{1}{2}\alpha_1 \cos \frac{1}{2}\alpha_{m-1}}$
m	$\frac{\sin \alpha_{1,m}}{kl}$	$\frac{\cos(\alpha_{1,m}-\frac{1}{2}\alpha_1)}{k \cos \frac{1}{2}\alpha_1}$	$\frac{\sin \alpha_{1,m}-k \sin \alpha_{1,m-1}}{\sin \alpha_m}$	$-\frac{l\{k \cos(\alpha_{1,m-1}-\frac{1}{2}\alpha_1)-\cos(\alpha_{1,m}-\frac{1}{2}\alpha_1)\}}{\cos \frac{1}{2}\alpha_1 \sin \alpha_m}$
$m+1$	$\frac{\sin \alpha_{1,m+1}}{kl}$	$\frac{\cos(\alpha_{1,m+1}-\frac{1}{2}\alpha_1)}{k \cos \frac{1}{2}\alpha_1}$	$\frac{k \cos(\alpha_{1,m+1}-\frac{1}{2}\alpha_{m+1})}{\cos \frac{1}{2}\alpha_{m+1}}$	$-\frac{kl \sin(\alpha_{1,m+1}-\frac{1}{2}\alpha_1-\frac{1}{2}\alpha_{m+1})}{\cos \frac{1}{2}\alpha_1 \cos \frac{1}{2}\alpha_{m+1}}$

All these values take the regular form when $k=1$.

If more alterations than one have taken place, or if a previous portion of the system does not belong entirely to a regular system, so that p is not unity before the diameter of the tube changes, it is only necessary to introduce the factors for all preceding changes into the expressions for the Gaussian constants and for the powers and separations of the lenses as they have been shown to occur in the present example.

§ (7) THE LIGHT PUPILS.—It might appear that a knowledge of the details of the leading lens system would be necessary to enable the properties suitable for later lenses to be determined, and hence necessary also in deciding what field of view can be included in a system of lenses to be inserted in a given tube when the number of lenses is fixed. This, however, is not the case, and a general view of what is physically attainable with an assigned number of lenses under these conditions may be gained without paying any regard to the lens system by which the result is to be achieved. To make this clear it is desirable to approach the problem from a different point of view from that hitherto adopted. For this purpose it is convenient to regard the problem as it will appear to a designer who is asked, without previous experience of periscope design, to devise a system to go into a given body and transmit light over a certain field of view, the magnifying power and the size of the central emergent pencil being assigned, while the illumination should be as uniform as possible, and in any case the marginal illumination must not be less than a known fraction of that at the centre.

These conditions are most simply represented by considering three circular apertures in the object space, one limiting the field which is to be visible, another the aperture of the objective situated at the instrument (the diameter being usually the diameter required for the emergent pencil multiplied by the magnifying power), and the last an aperture, usually situated between the other two, which cuts down the

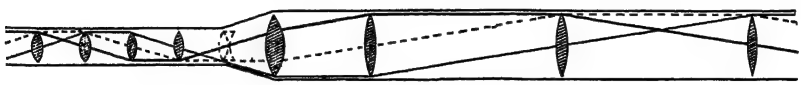


FIG. 31.

areas of marginal pencils without affecting those at the centre of the field. Alternatively corresponding conditions may be given for the image space. The three apertures determine a region of space containing all the useful light rays, and the instrument should have its optical elements of such sizes and in such positions that no ray which does not cross the boundary of this space is obstructed in its passage from the object through the instrument to the observer's eye. It is assumed that the apertures are circles lying in planes normal to the axis of symmetry of the instrument, and that their magnitudes and positions are known for either the object space or the image space. In the illustrations to be given it will be assumed that the various lenses between the objective and the eye lens may be permitted to have an aperture not exceeding a common value d. The modification necessary when the maximum aperture varies from lens to lens will be obvious.

Suppose, then, that the boundary to the channel containing the rays to be transmitted by the instrument is defined by means of three circular apertures which are met by a plane through the axis in P, P′; R, R′; S, S′ (*Fig.* 32). The section of the boundary by this plane is the convex hexagon RSP′R′S′P, the sides SP′ and S′P passing through infinity. The image of this hexagon formed by any part of the optical system of the instrument

will likewise be a convex hexagon, and this image will determine the channel for the refracted rays. After every refraction there is necessarily some part of the boundary which extends to infinity, and this may correspond to any assigned normal section of the original channel.

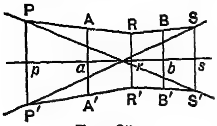


Fig. 32.

Only that part of the boundary which lies between successive lenses is of interest from the present point of view. These various segments pieced together and regarded as a whole constitute repeated outlines of the whole hexagon, which always presents its convex side to the axis, but the outline has an additional angle wherever a lens is encountered. At such places one section ends and another begins, and the angle at their junction is opposite in sign to the angles of the hexagon if the lens is positive. Lenses are introduced wherever necessary to bend back the boundary of the channel from encroaching over the prescribed bounds.

Fig. 33 illustrates the effects that may be produced by successive refractions, the cross

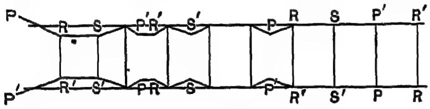


Fig. 33.

lines representing lenses. The various points of the original hexagon are identified by the same letters in their refracted images. The only limit to the amount of the initial hexagon that may be imaged between successive lenses arises from the fact that a pair of opposite sides of the hexagon always extend to infinity, and it is not possible in consequence to image one half of the outline on one side of the boundary within a single cell of the body.

The separation between the lenses limiting a given cell is readily found. Suppose that the lenses meet the boundary in planes corresponding to AA′ and BB′ of *Fig.* 32. If each of the transverse lines AA′ and BB′ has its image of diameter d, the transverse magnifications are d/AA' and d/BB'. By a well-known elementary theorem the longitudinal magnification will be the product

$$\frac{d}{\mathrm{AA}'}\cdot\frac{d}{\mathrm{BB}'},$$

and the separation of the lenses will thus be

$$\frac{d^2\,.\,ab}{\mathrm{AA}'\,.\,\mathrm{BB}'},$$

where a and b are the axial points of the planes AA′ and BB′. The total length between the extreme lenses after any number of refractions is found by adding together the lengths corresponding to the successive intervals. The law already derived from different considerations, that this length will be proportional to the square of the diameter of the tube and to the number of lenses, and inversely proportional to the diameters of the field of view and of the external aperture of the system, at once follows.

§ (8) Standard Cases. (i.) *General Spacing.*—Particular cases which require special attention occur when the lenses occupy positions corresponding to the limiting apertures PP′, RR′, SS′. If a lens is placed at every aperture image the greatest length is attained when the hexagon is developed into a straight line, so that the rays occupy the whole available space within the body of the instrument. The length reached is given by the sum

$$d^2\left\{\frac{sp}{\mathrm{SS}'\,.\,\mathrm{P'P}}+\frac{pr}{\mathrm{P'P}\,.\,\mathrm{R'R}}+\frac{rs}{\mathrm{R'R}\,.\,\mathrm{S'S}}+\frac{sp}{\mathrm{S'S}\,.\,\mathrm{PP}'}+\frac{pr}{\mathrm{PP}'\,.\,\mathrm{RR}'}+\ldots\right\},\qquad(12)$$

where it is assumed that SS′ corresponds to the position of the first lens. All the terms in this sum necessarily have the same sign, as changes of sign take place simultaneously in the numerator and the denominator. Several alternative arrangements suggest themselves. Suppose, for instance, that a lens is placed at the image of every alternate aperture only. The length is then

$$d^2\left\{\frac{sr}{\mathrm{SS}'\,.\,\mathrm{R'R}}+\frac{rp}{\mathrm{R'R}\,.\,\mathrm{PP}'}+\frac{ps}{\mathrm{PP}'\,.\,\mathrm{SS}'}+\ldots\right\},\qquad(13)$$

the terms having exactly the same values as before but differing in order. If in these cases the total number of lenses is a multiple of three the lengths of the two instruments will be exactly equal.

As another example, suppose there were no field lenses, so that no lens is placed at any image of the aperture PP′. The sum is now

$$d^2\left\{\frac{sr}{\mathrm{SS}'\,.\,\mathrm{R'R}}+\frac{rs}{\mathrm{R'R}\,.\,\mathrm{SS}'}+\frac{sr}{\mathrm{S'S}\,.\,\mathrm{RR}'}+\ldots\right\},\qquad(14)$$

every term being equal to one of the terms in (12). The other apertures may be treated in the same way, and it is evident that the first and second methods of construction will not give either the longest or the most compact instruments if the values of the individual terms differ. The estimation of the relative values of the terms is thus interesting, and this

can be done by a very simple construction. Let pS and Ps meet RR' in t and u respectively (*Fig.* 34).

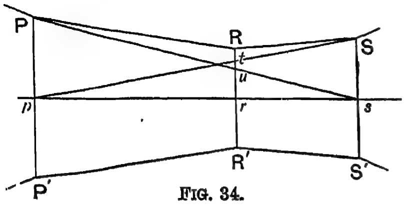


FIG. 34.

The relative values are in the ratios of the products

$$ps\,.\,RR' : pr\,.\,SS' : rs\,.\,PP',$$

that is, of the triangles

$$pRs : pSr : rPs,$$

which are equal to the triangles

$$pRs : pts : pus,$$

which are in the ratio of their heights

$$Rr : tr : ur.$$

The length (12) may therefore be written as

$$\frac{d^2\,.\,ps}{PP'\,.\,SS'}\left\{1+\frac{tr}{Rr}+\frac{ur}{Rr}+1+\frac{tr}{Rr}+\frac{ur}{Rr}+\ldots\right\}. \quad (15)$$

PP' and SS' may be taken to represent the actual object and the entrance pupil, so that the factor outside the bracket will be known. The greatest attainable length of any cell not bounded at an image of the aperture which cuts down the marginal light has then a definite value. The next term gives the length of a cell bounded by an image of the field and of the intermediate stop. As ps may be always considered large in comparison with rs, the utmost that can be done is to make this cell equal in length to the previous one, by making the diameter of the intermediate stop image of the same size as the entrance pupil. This involves an immediate fall in the light intensity as soon as the object departs from the axis. The length of the third cell, which is bounded by images of the intermediate stop and of the entrance pupil, depends upon the relative illumination at the edge and at the centre of the field of view. If the width of the marginal beam in the axial plane of symmetry is half that of the central beam, PRs will be a straight line and the lengths of the first and third cells will be equal. If the edge illumination is increased, the length of this cell is diminished, and if the illumination falls, the length may be increased. When the width of the marginal beam is greater than half the central width, it will evidently be possible to attain a greater length per lens when lenses at images of RR' and SS' do not immediately succeed one another.

(ii.) *Uniform Spacing.*—The case when all three cells are of equal length is a convenient one to take as a standard of reference, and systems of this type possess great flexibility. It is always possible to attain a given standard extent of field with a standard aperture and the distribution of illumination characteristic of the type with any given number of cells. The length per cell thus gives complete information about such systems. This may be illustrated by the following examples. The cells begin with the first objective, indicated by S, and end with the formation of a real upright image of the object, denoted by P. If the system ended with P′ the final image would be inverted.

TABLE IV

No. of Cells.	Lens Positions.	Remarks.
2	S, R′, P	Not divisible into separate telescopes.
3	S, P′, S′, P	Not divisible into separate telescopes.
3	S, P′, R′, P	Not divisible into separate telescopes.
3	S, R′, S′, P	No field lens.
4	S, P′, R′, S′, P	
5	S, R′, P, S, R′, P	Not divisible into separate telescopes.
6	15 solutions	A lens coincides with at least one image of each stop.
7	35 solutions	At least two lenses coincide with images of each stop.

In general with $2n$ inversions and p cells the number of solutions is

$$\frac{p!}{(6n-p-2)!\,(2p+2-6n)!},$$

where p may have any value from $3n-1$ to $6n-2$.

It does not follow that because several solutions enable the same length of body to be obtained with the same number of lenses that these are all equally desirable. Other things being equally satisfactory, the solution preferred will be that which involves the use of the least powerful lenses. It should be

immediately evident that the omission to place a lens where an image of a certain aperture is formed will involve increases in the powers of the lenses at the two neighbouring images, since the deviations at the margins of these lenses must be greater. The failure, therefore, to fill completely the space available for the rays will be associated with increased lens powers, and consequently with greater trouble due to secondary colour effects and increased curvature of field. To determine how serious these defects are likely to be, the powers of the lenses are required. There is no need to consider the refracted light to determine these powers, but attention may be directed as before to the object space hexagon. If, for example, successive lenses are to be located at images of PP′, RR′, SS′, the lens at the image of RR′ is required to deviate the image of RS into alignment with the preceding image of PR. Now the angular magnification at any plane is inversely proportional to the linear magnification, and the angular deviation is equal to the product of the lens power and the distance from the axis at which refraction takes place. It at once follows that the power of the lens required is $(\mathrm{RR}'/d)^2$ times the power of the lens which would produce the corresponding effect in the object space boundary. This is in reality the same result that has already been reached regarding the length of the instrument, since the focal lengths of the lenses evidently bear a fixed relation to the lengths of the cells, but it serves to show how great is the advantage from every point of view in making the diameter of the body as great as possible. Various expressions are easily found for the power of the lens necessary at RR′. Let this be denoted by $\kappa(d/\mathrm{RR}')^2$, so that κ is the power of the lens required in the instrument. The deviation at R is then $\frac{1}{2}(\kappa d^2/\mathrm{RR}')$, and the area of the triangle PRS by various expressions leads to the equations, when inclinations to the axis are assumed to be small,

$$\frac{\kappa d^2}{\mathrm{RR}'} \cdot pr \cdot rs = 2\rho\mathrm{R} \cdot ps \quad . \quad . \quad (16)$$

$$= \mathrm{PP}' \cdot rs + \mathrm{RR}' \cdot sp + \mathrm{SS}' \cdot pr, \quad . \quad (17)$$

where ρ is the point of intersection of $\mathrm{R}r$ and PS. Equation (17) itself shows the result reached above in a different way that κd^2 is unaltered by refraction. For it may be rewritten in the form

$$\kappa d^2 = \frac{\dfrac{rs}{\mathrm{RR}' \cdot \mathrm{SS}'} + \dfrac{sp}{\mathrm{SS}' \cdot \mathrm{PP}'} + \dfrac{pr}{\mathrm{PP}' \cdot \mathrm{RR}'}}{\dfrac{pr}{\mathrm{PP}' \cdot \mathrm{RR}'} \cdot \dfrac{rs}{\mathrm{RR}' \cdot \mathrm{SS}'}},$$

and each term on the right is invariant on refraction. The expressions of this type for the powers of the lenses at images of PP′, RR′, and SS′ when no images are without lenses are evidently

$$\left.\begin{aligned} d^2\kappa_{\mathrm{RSP}'} &= d^2\kappa_{\mathrm{R}'\mathrm{S}'\mathrm{P}} \\ &= \mathrm{SS}'\frac{\mathrm{RR}' \cdot sp + \mathrm{SS}' \cdot pr + \mathrm{P}'\mathrm{P} \cdot rs}{rs \cdot sp} \\ d^2\kappa_{\mathrm{S}'\mathrm{PR}} &= d^2\kappa_{\mathrm{SP}'\mathrm{R}'} \\ &= \mathrm{P}'\mathrm{P}\frac{\mathrm{SS}' \cdot pr + \mathrm{P}'\mathrm{P} \cdot rs + \mathrm{R}'\mathrm{R} \cdot sp}{sp \cdot pr} \\ d^2\kappa_{\mathrm{PRS}} &= d^2\kappa_{\mathrm{P}'\mathrm{R}'\mathrm{S}'} \\ &= \mathrm{RR}'\frac{\mathrm{PP}' \cdot rs + \mathrm{RR}' \cdot sp + \mathrm{SS}' \cdot pr}{pr \cdot rs} \end{aligned}\right\}, \quad (18)$$

where all the factors are considered with due regard to sign. In each numerator two terms have one sign and the third the opposite sign, in each case giving rise to a positive κ. In the standard case the terms are all equal, and the focal length of each lens is equal to the length of a cell. This is readily seen geometrically by considering the intercepts on the third aperture by the line forming the boundary between the other two.

The cases in which an image of an aperture is without a lens may be treated by analogy with (18),

$$\left.\begin{aligned} d^2\kappa_{\mathrm{RSR}'} &= d^2\kappa_{\mathrm{R}'\mathrm{S}'\mathrm{R}} = d^2\kappa_{\mathrm{S}'\mathrm{RS}} \\ &= d^2\kappa_{\mathrm{SR}'\mathrm{S}'} = \frac{2\mathrm{RR}' \cdot \mathrm{SS}'}{rs} \\ d^2\kappa_{\mathrm{SP}'\mathrm{S}'} &= d^2\kappa_{\mathrm{S}'\mathrm{PS}} = d^2\kappa_{\mathrm{PSP}'} \\ &= d^2\kappa_{\mathrm{P}'\mathrm{S}'\mathrm{P}} = \frac{2\mathrm{PP}' \cdot \mathrm{SS}'}{ps} \\ d^2\kappa_{\mathrm{PRP}'} &= d^2\kappa_{\mathrm{P}'\mathrm{R}'\mathrm{P}} = d^2\kappa_{\mathrm{R}'\mathrm{PR}} \\ &= d^2\kappa_{\mathrm{RP}'\mathrm{R}'} = \frac{2\mathrm{PP}' \cdot \mathrm{RR}'}{pr} \end{aligned}\right\}, \quad (19)$$

showing that in the standard case the omission of a lens requires the power of each neighbouring lens to be doubled. Again, if a lens is omitted on each side of a given lens expressions are obtained corresponding to (18), but with every term of the same sign, so that in the standard case the power of the lens is three times its previous value. The expressions are

$$\left.\begin{aligned} d^2\kappa_{\mathrm{PSR}'} &= d^2\kappa_{\mathrm{P}'\mathrm{S}'\mathrm{R}} \\ &= \mathrm{SS}'\frac{\mathrm{PP}' \cdot sr + \mathrm{SS}' \cdot rp + \mathrm{R}'\mathrm{R} \cdot ps}{rs \cdot sp} \\ d^2\kappa_{\mathrm{R}'\mathrm{PS}} &= d^2\kappa_{\mathrm{RP}'\mathrm{S}'} \\ &= \mathrm{PP}'\frac{\mathrm{R}'\mathrm{R} \cdot ps + \mathrm{PP}' \cdot sr + \mathrm{SS}' \cdot rp}{sp \cdot pr} \\ d^2\kappa_{\mathrm{SR}'\mathrm{P}} &= d^2\kappa_{\mathrm{S}'\mathrm{RP}'} \\ &= \mathrm{R}'\mathrm{R}\frac{\mathrm{SS}' \cdot rp + \mathrm{R}'\mathrm{R} \cdot ps + \mathrm{PP}' \cdot sr}{pr \cdot rs} \end{aligned}\right\}. \quad (20)$$

By considering the various cases that may arise it is seen that the effect of omitting to place a lens at an aperture image is to contribute for this vacant aperture twice the normal contribution to the sum of the powers of the lenses, assuming that the standard case is under consideration. If account is taken of a final field lens but not of an eye lens, with $2n$ inversions and p cells the sum of the powers will be proportional to $12n-p-3$. In this respect all arrangements which involve the same number of lenses and the same number of inversions are on an equality in the standard case.

In the general case the increase in power associated with the omission of a lens at an image of P is

$$-\frac{(SS'.pr+P'P.rs+R'R.sp)^2}{d^2.pr.rs.sp}, \qquad (21)$$

the term in the bracket being that in the numerator of the fraction in the second of equations (18). The corresponding expressions for other cases are easily written down. The sum of the additions incurred by each separate omission gives the total addition to be made to the sum of the powers.

A convenient expression for the increment in power in terms of the lengths of the cells is

$$\frac{\kappa^2 l_1 l_2}{l_3}, \quad . \quad . \quad . \quad . \quad (22)$$

where κ is the power of the lens which it is proposed to omit, l_1 and l_2 are the lengths of the cells on either side, and l_3 is the length of the combined cell when the omission is made. The three lengths l_1, l_2, l_3 are those considered earlier in this discussion when the length attainable by a periscope was investigated.

It is evident that when weight is attached both to the attainment of great length and to a minimum value of the power sum, the most favourable type of construction is that which has been generally adopted for submarine periscopes, with a lens at every image of the three apertures.

(iii.) *Advantage of Non-uniform Spacing.*—In investigating the attainment of great length attention was directed to each cell as a separate element. It has still to be considered whether, when the conditions permit of it, any advantage will be gained by deliberately introducing inequalities in the lengths of the three cells. Suppose the lengths of the cells exceed the mean value L by amounts λ_1, λ_2, λ_3, where $\lambda_1+\lambda_2+\lambda_3=0$. From (18)

$$\kappa_1=\frac{l_2+l_3-l_1}{l_2 l_3},$$

with symmetrical expressions for κ_2 and κ_3, so that

$$\kappa_1+\kappa_2+\kappa_3 =\frac{3L^2-2(\lambda_1^2+\lambda_2^2+\lambda_3^2)}{L^3-\frac{1}{2}L(\lambda_1^2+\lambda_2^2+\lambda_3^2)+\lambda_1\lambda_2\lambda_3}$$

$$=\frac{3}{L}-\frac{\lambda_1^2+\lambda_2^2+\lambda_3^2}{2L^3-L(\lambda_1^2+\lambda_2^2+\lambda_3^2)} \text{ approximately,}$$

showing that inequalities in length tend to reduce the aberrations which depend on the sum of the powers. In this respect the standard form is the least satisfactory solution of its class, but the advantage to be gained by uneven spacing is small, since it is a second order effect. For instance, if the lengths are in the ratios 2 : 3 : 4, the reduction in the power sum is slightly more than 4 per cent.

§ (9) Separation of Image and Field Lenses.—It has hitherto been assumed that the lenses will be placed in the planes at which aperture images are formed. In the case of field lenses exact coincidence with the image is generally undesirable, and the effect of placing the lenses in other planes requires consideration. Obviously the sizes of the lenses are determined by the sections of the channel to which their positions correspond. Suppose in *Fig.* 32 the planes AA′ and BB′ are proposed as lens planes. The relation between transverse and longitudinal magnitudes is

$$AA'.pr=PP'.ar+RR'.pa,$$

and thus if the apertures PP′, RR′, SS′ relate to a standard form,

$$AA'.SS'.pr=(ar.ps+pa.rs)\frac{d^2}{l}$$

$$=as.pr.\frac{d^2}{l},$$

showing that a cell bounded by AA′ and SS′ will be of standard length. Similarly

$$PP'.BB'=pb.\frac{d^2}{l},$$

and combining the two cases,

$$AA'.BB'.ps=pb.as.\frac{d^2}{l}.$$

Thus if l' is the length of the cell bounded by AA′ and BB′,

$$\frac{l'}{l}=\frac{ab.ps}{pb.as}$$

or

$$\frac{l-l'}{l}=\frac{pa.bs}{pb.as},$$

showing that if two successive lenses are not in planes corresponding to the angular points of the hexagon the length of the cell they bound is reduced.

§ (10) Completion of the Instrument.—The theory developed in the foregoing sections suffices to enable the special problem presented

by the submarine periscope to be solved without difficulty in any given case. When the instrument has only a single magnifying power the arrangement of lenses for the main part of the tube may first be investigated, the objective may be arranged to fit in with this system, and finally the eyepiece be added to secure a suitable power. The problems here involved do not differ essentially from those presented by any simple telescope.

§ (11) Bi-focal Instruments. — Though very great importance is attached in submarine periscopes to the attainment of a very wide field at a low power, it is also advantageous to be able at will to substitute a higher magnifying power with a correspondingly smaller field of view. In small telescopes (*q.v.*) this is usually secured by the use of a variable power eyepiece, but for a submarine periscope this method of changing the power is unsatisfactory. It has to be remembered that the brightness of the image is a factor of outstanding importance whether the power is high or low, and therefore at the high power it is necessary for the pupil to be filled with light. If the diameter of the eye pupil for which the instrument is designed is e and the apparent field of view is Ω, the minimum diameter of the beam of light which must be carried through the instrument is m_2e and the minimum angular field of view to be carried to the eyepiece is Ω/m_1, where m_1 is the magnifying power for the low power and m_2 for the high power. The duty to be performed by the main part of the instrument is thus represented approximately by

$$\frac{e\Omega m_2}{m_1}$$

when a variable power eyepiece is used. On the other hand, when the power is changed by varying the objective the corresponding expression is $e\Omega$ whatever m may be. As m_2 is usually about four times m_1, it follows that the results attainable with a variable power objective are approximately four times as good as can be obtained from a variable power eyepiece.

As with a telescope, it is essential that the change of power shall not necessitate any change in the adjustment for focussing the eyepiece on the image. Accordingly, when the power has been changed the image formed by the objective must remain in the same plane as it originally occupied; the consequent portions of the system will then transmit the light under conditions exactly similar for both powers. The discussion of variable power eyepieces [1] shows that a continuous variation of power without a change of image position is possible if two lenses move suitably relatively to the rest of the system. Arrangements of this kind are considered unsuitable in submarine periscopes, for they not only involve a considerable enlargement of the outer tube for the accommodation of the necessary mechanism where it is particularly desired to reduce its diameter as much as possible, but also the likelihood of their failure to act through distortion or slight damage to the periscope renders their introduction distinctly undesirable in so vital an instrument.

The exclusion of these systems which give a continuous variation of power implies that the objective is rapidly exchanged for another of different focal length, the object under examination being momentarily obscured while the change is being effected. This change has been secured in a variety of ways. In an instrument due to Messrs. Ross, separate fixed objectives are used for the high and the low powers, and one or the other is obscured by shutters. This arrangement involves a slight loss of light at the high power, since a reflecting prism has to be slightly cut away to allow the low power beam to pass. A design by Messrs. Zeiss makes use of the fact that with any lens the separation of the object and the image is the same whether the magnification is m or $1/m$. Thus, if the ratio of the two magnifying powers is 4 : 1, values for m of 2 and $\frac{1}{2}$ would be selected; if the moving lens is of focal length f and a preliminary lens presents to it as an object an image distant about $\frac{3}{2}f$, this second lens will form an image twice the size at a distance of approximately $3f$ (*Fig.* 35). If now the second lens is

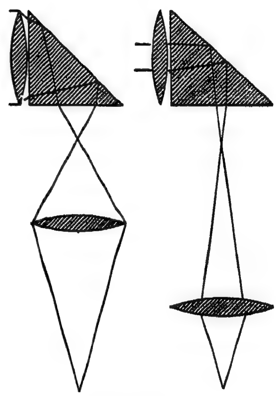


Fig. 35.

moved a distance $\frac{3}{2}f$ further away from the first lens, the resulting second image will be half the size of the first or one-quarter the size of the original second image, but its position in the instrument will be unchanged. The movement of a single lens along a tube in this way is not open to the same objections as in

[1] See "Eyepieces."

the more complicated movements of the variable power eyepiece type. The chief objection to this system is that unless the instrument can be corrected for aberration when this lens is of symmetrical construction the lens should, in addition to its translation, be reversed on moving between its two positions.

An entirely different system is one due to Sir Howard Grubb (*Fig.* 36). In this case two

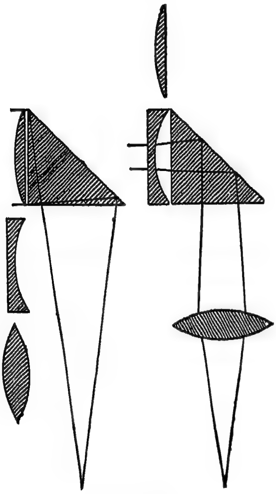


FIG. 36.

objectives are mounted in the same frame which slides in front of the entrance window, so that either lens may receive the light. One of these lenses acts as an objective alone for one power. For the other power the second lens occupies the position of the first, the frame being moved parallel to the tube axis through the distance separating the lens centres. It is evidently necessary that this lens shall differ in power from the first and that another lens shall be introduced into the system, in order that the power shall be different and the image in the same position. In the arrangement used this supplementary lens moves about a hinge near one side of the tube, and lies close to the tube when the single objective is used, but turns through a right angle into the optical system when the second power is employed. By a simple but ingenious mechanism a single movement effects both changes. If t_1 is the separation of the lenses when the double objective is used, and t_2 is the distance from the second of these lenses to the image, so that t_1+t_2 is with sufficient approximation the distance from the single objective to the image plane, the powers of the lenses must be for the single objective

$$\kappa=\frac{1}{t_1+t_2},$$

and for the double objective

$$\kappa_1=\frac{m}{t_1+t_2}-\frac{m-1}{t_1},$$

$$\kappa_2=\frac{m-1}{m}\left(\frac{1}{t_1}+\frac{1}{t_2}\right),$$

giving a power for the combination of

$$\kappa_{1,2}=\frac{m}{t_1+t_2}.$$

A particularly attractive means of changing the power is that introduced by Messrs. Goerz (*Fig.* 37), in which use is made of the fact that an objective composed of a positive and a negative lens well separated from one another have their principal planes very unsymmetrically situated. Under suitable conditions the mid-point of the actual lens system may approximately coincide with the point midway between the object and image planes when the magnification differs appreciably from unity. The system is thus a special case of the one utilised by Messrs. Zeiss, but the translation is converted into a rotation. The arrangement actually adopted is to use two pairs of positive and negative lenses mounted as a whole with two right-angled deflecting prisms. This complete block is rotated

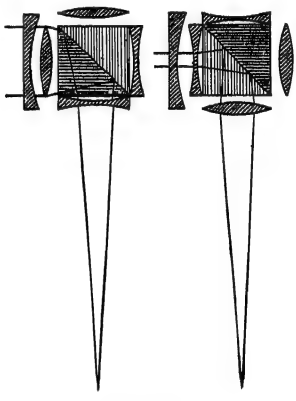


FIG. 37.

through two right angles about an axis normal both to the vertical tube axis and to the horizontal axis of the incident light. The use of two separate objective systems may be due to difficulties in connection with the removal of aberrations. This system must be preceded by a fixed negative lens to yield a virtual image at the same distance from the axis of revolution as the first image formed in the periscope tube. If t is the separation between the lenses and d is the distance of both object and image from

the nearest lens, the system is determined from

$$\kappa_1=\frac{1}{d}+\frac{1+m}{mt},$$

$$\kappa_2=\frac{1}{d}+\frac{1+m}{t},$$

$$\kappa_{1,2}=-\frac{t}{d^2}-\left(m+\frac{1}{m}\right)\frac{1}{d},$$

where m of course is negative.

The most usual powers of bi-focal periscopes are $1\frac{1}{2}$ and 6, the former rather than power 1 giving an impression equivalent to a naked-eye view when the apparent field of view is limited. The value adopted for e is usually about 6 mm. and the apparent field may exceed 60°. In a 25-ft. periscope the narrow portion may be 4 ft. in length, and the external diameters will be approximately 6 in. and $2\frac{1}{2}$ in. for the main and top tubes respectively.

§ (12) Sky-searching Periscope.—The earlier periscopes provided only for a view towards the horizon, and all light entered through a fixed window in the side of the tube. It became necessary at a later date to provide in addition means for searching the sky from the horizon to beyond the zenith. The angle here involved is too great to make it practicable to include the whole in a system with a fixed axis. Apart from this, since the definition is only at its best near the axis, it is necessary to be able to secure at will a horizontal direction for the axis. A movable reflector accordingly became necessary at the head of the instrument, and the most satisfactory arrangement consists of two isosceles prisms silvered on their bases, which are cemented together into a frame with their bases parallel (*Fig.* 38).

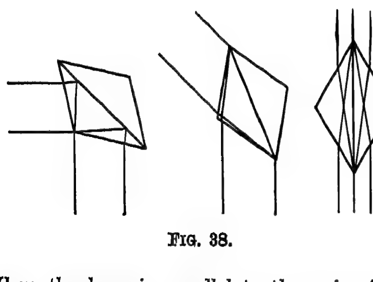


Fig. 38.

When the base is parallel to the axis of the instrument an axial beam of light continues in the same direction after reflection, and, as the inclination of the base to the tube axis is altered by rotation about an axis perpendicular to the length of the tube, the deviation of the light is altered by twice this amount. The use of two prisms mounted in this way gives more equal illumination for prisms which can rotate within a given space than is possible when only a single prism is used. The window through which light first enters the instrument, and which is chiefly provided to prevent the entrance of water into the periscope, is in this case inclined to the axis of the instrument at about 45° instead of being parallel thereto as in earlier instruments. Supplementary systems of prisms may be used to secure more advantageous use of incoming light.

§ (13) Miscellaneous.—The foregoing description is confined to the optical systems which have been actually employed in periscopes. A great variety of interesting instruments have been designed, offering, for instance, panoramic views of the whole horizon arranged in a circular view with an enlarged direct view in the middle, but it is impossible in a single article to attempt to describe them. Actual periscopes also must be provided with much auxiliary equipment in order to remain serviceable under the conditions of use. In some cases, they are from time to time filled with dry compressed air to render the ingress of water more difficult and to prevent the deposition of moisture on any internal surface when the temperature of the sea is low. The discussion of all these necessary but non-optical auxiliary features has not been attempted in this article.

§ (14) Illustrations of Modern Periscopes.—It will be of interest to add some illustrations of recent periscopes showing the actual appearance of the two ends of the instruments. Exhaustive illustration of the many variations is impossible, but those that are given here will serve not only to give an idea of the appearance of the instruments in actual use, but will also indicate in a few directions the developments of which the periscope is susceptible. These illustrations are from photographs of periscopes made by Messrs. Barr & Stroud, by whose courtesy it has been possible to include the illustrations and descriptions of this and the following sections. Incidentally the principle of the arrangement for securing two magnifying powers which this firm has adopted, and which differs from those illustrated in § (11), is shown diagrammatically in *Fig.* 42.

It is not practicable within the limits of space imposed by the pages of this dictionary to show either a general outline of ordinary periscopes or illustrations of the internal mechanism. It may, however, be said that the length of submarine periscopes varies from very short instruments for night use which do not greatly exceed 3 metres in length to instruments reaching 15 metres or more in length. In many of the recent instruments the reduction in the diameter of the top section, which in early instruments was a very pronounced

feature, has been much more slight or even entirely absent. The reason for this is that the instruments are used in a different way. In approaching the object of attack the small-topped periscopes were kept for some considerable time above the surface of the sea. Though the periscope itself might escape observation, it was found that the presence of the submarine was detectable through the track it made in the water when used in this way. The more general method now is to elevate the periscope above the surface at intervals for a few seconds only at a time. Under these conditions, even with a thick top, the periscope usually runs little risk of being seen, and the surface disturbance created when this method of operation is adopted resembles very closely the normal surface irregularities, and so is of no assistance in detecting the approach of an enemy. The possibility of using the thicker head with lenses of larger diameter is of very special importance for very long periscopes, since it is in these instruments that the loss of light due to the large number of lenses becomes most serious, and the narrowness of the smallest section is the chief factor, as has been shown earlier, in increasing the number of lenses that are necessary. The raising and lowering of the periscope is usually performed by hydraulic or electrical power.

The material of which the main tubes are made is of much importance. A periscope made of ordinary steel would inevitably be strongly magnetic, and would affect the reliability of the indications of a magnetic compass in a quite intolerable way. The alternatives are to use a bronze or a non-magnetic steel. In the early periscopes a bronze tube was employed, but later this was replaced by a special nickel steel. At the best, however, tubes of the latter material are somewhat magnetic, and in addition they are slightly corroded by the action of sea-water. This corrosion is of more importance than might at first sight appear since it may cause the tube to work stiffly in the long gland where it pierces the hull of the boat, and so cause difficulty in manipulation. In both of these respects a better material is available, and Messrs. Barr & Stroud prefer to utilise Immadium bronze. This is completely non-magnetic and possesses remarkable powers of resisting corrosion in sea-water, its surface retaining its original polish after prolonged immersion. Needless to say, it possesses the mechanical qualities necessary in a material which must withstand very severe stresses without any considerable deflection.

Fig. 39 shows a photograph of the upper end of a periscope of the sky-searching type. The glass which comes in contact with the water is a window with flat surfaces. Inside may be seen the prism by the rotation of which about a horizontal axis the direction of sight may be varied from the zenith to below the horizon, according to the principle illustrated in *Fig.* 38.

The lower end of the periscope is illustrated in *Fig.* 40. The handles are used to rotate the instrument about its axis, and so enable the line of sight to be directed to any part of the horizon as required. The bearing of any object to which the periscope is directed may be read on the scale seen in the upper part of this figure. When the instrument is not in use the handles may be folded into a vertical position out of the way. The eyepiece, surrounded by a rubber eye-guard, is seen between and slightly above the handles. The kind of view obtained

FIG. 39.

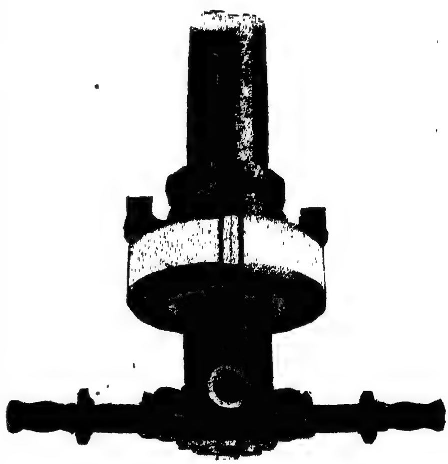


FIG. 40.

on looking into the eyepiece is shown in *Fig.* 41. In focus together with the image of distant external objects, the graduations of vertical and horizontal scales are seen. These correspond to suitable angular intervals by means of which the apparent sizes of objects may be more accurately estimated. The need for such measurements will be fully realised when the more precise measuring devices described below are considered.

Inset in the top of the field of view, so as to be visible without taking the eye from the eyepiece, is a view of a portion of the bearings scale. This is obtained by the introduction of a small auxiliary prismatic telescope.

In addition to reading the bearing, it is obviously of importance that any desired object should be readily brought to the centre

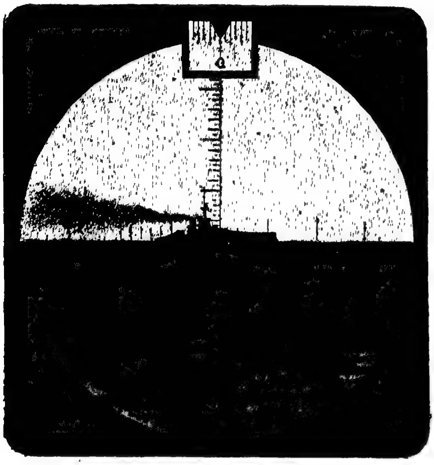


FIG. 41.

of the field under the conditions in which it can be best seen, without the observer being under the necessity of removing his eye from the eyepiece. This evidently involves suitable control over the angle of elevation of the line of sight in addition to its bearing, and ability to adjust the focus, and in special cases to interpose or remove rapidly suitably coloured screens. In this particular periscope the top prism may be adjusted for elevation as required by rotating one of the bearing handles about its axis, and the focussing adjustment and the light filters are controlled by means of the two milled heads seen projecting from the bottom of the periscope in *Fig.* 40. It will be recognised that all these controls are placed very conveniently for the observer. The coloured glasses are placed within the tube and lie close to the eye lens. The focussing is effected by changing the position in the tube of a movable lens. The principles on which the optical system is constructed are shown in *Fig.* 42, which is necessarily diagrammatic and not to scale. The eyepiece consists of lenses M and N, of which the former is cemented to the lower prism L to reduce the number of glass to air surfaces, and thus to obviate an avoidable loss in the amount of light transmitted by the instrument. Focussing is accomplished by movement along the axis of the collector lens K. This lens, together with the eyepiece and the objective H, forms an astronomical telescope. To transmit a sufficient quantity of light down the length of the periscope and secure a suitable field of view, this lower telescope is preceded by another astronomical telescope composed of three lenses E, F, G. Of these E corresponds to the eyepiece, F is a field lens, and G may be regarded as the objective. Considered in this way this telescope produces an angular magnification measured by the ratio of the focal lengths of lenses G and E. If E is regarded as the objective and G as the eye lens, the way this telescope is actually used, it appears to diminish objects viewed through it. The combination of the two telescopes of course is itself a telescope, and as EFG produces an inverted image, and this is subsequently reinverted when viewed through the second telescope HKMN, the combination yields an upright image. To make the system suitable for use as a periscope it is only necessary to insert an upper prism B, and before this to place a plane surfaced window A to prevent the ingress of water into the instrument. This, in fact, constitutes a high-power telescope, the magnifying power being usually 6 with a field of view of 10°. A lower power is obtained by the very simple device of introducing a reversed Galilean telescope in front of the high-power system. The negative lens of this

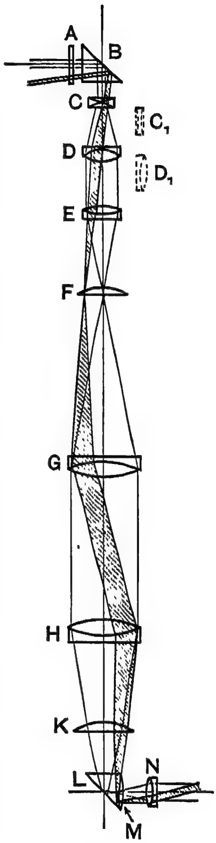


FIG. 42.

telescope is marked C and the positive D. When the high power is wanted these are swung to one side of the tube as at C_1 and D_1. The Galilean telescope is ordinarily constructed to be of magnifying power 4 when used alone as a telescope, with C as the eye lens and D as the object glass. As used in the periscope the magnifying power is thus $\frac{1}{4}$, making the power for the complete low-power system $1\frac{1}{2}$. The field of view is increased in the reciprocal ratio, and is thus approximately 40°. The diagram illustrates the passage of a central and of an oblique beam of rays through the low-power system, and the reduction of the area of the oblique beam in comparison with the central beam is clearly seen. This reduction may not

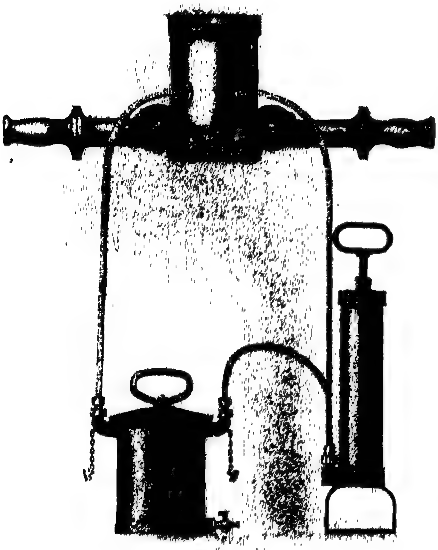


FIG. 43.

be carried beyond the point at which the difference in intensities at the centre and margin becomes apparent to the eye when the change takes place continuously from centre to margin.

When the high-power system is substituted the paths of the rays between D and N are not affected, but in the upper portion of the instrument the paths are modified and may be found by tracing back the paths between D and E. The variation in the illumination from centre to margin is substantially the same for both powers, for this depends chiefly on the separation of the two objectives G and H. The change of power is effected by rotating one of the training handles shown in *Fig.* 40; the other, as has been mentioned, is used to change the elevation of the line of sight. In general the low-power telescope is employed, and a change is only made to the high power when it is desired to observe an object, that has been sighted, more minutely than is possible with the low power. By using the low power, which embraces an angular area sixteen times as great as the high power, large regions can be observed much more rapidly than with the high power with its smaller field of view.

As has been mentioned, it is necessary from time to time to dry the air within the periscope. For this purpose a desiccating chamber is provided in which the damp air from the bottom of the periscope is dried by passing over calcium chloride prior to being forced by a pump to the top of the instrument. The connections to the desiccator are made at the back of the lower end of the periscope as shown in *Fig.* 43. When the desiccator is not in use the inlet and outlet are protected by water-tight plugs.

§ (15) RANGE, COURSE, AND SPEED MEASUREMENTS.—The periscope as described above fulfils functions equivalent to the look-out, with or without the aid of a telescope, on a surface craft. The surface vessel in addition makes use of a number of other optical instruments, such as gun-sights and range-finders. Any corresponding instruments for the submarine must assume a periscopic character, and as the power of making accurate measurements is as important for the under-water as for the above-water ship, it is natural to find that attempts have been made to obtain such measurements, in spite of the special difficulties inseparable from the peculiar conditions of observation from a submerged boat.

For finding the distance of a visible object, one of the simplest methods is to measure the angle it subtends when the dimensions of the object are known. The graticule seen in *Fig.* 41, the divisions of which have known values, enables an approximate value of the range to be found. The method can be increased in accuracy by using a variable wedge to produce a double image by placing it in front of a portion of the beam. A convenient arrangement is one in which the wedges are annular in form, as shown in *Fig.* 44. When the thickest portion of one wedge is superposed on the thinnest portion of the other and *vice versa*, the two are equivalent to a plane sheet of glass if of equal angle, and the images formed by light which has traversed the wedges are not displaced to one side of those formed by the central portion of the pencil. If one prism is rotated in one direction and the other turned through an equal angle in the other direction, the outer rays will be displaced in the plane perpendicular to that originally containing the lines of greatest slope, and this displacement receives its maximum value when each prism has been turned through one right angle, the

position illustrated in *Fig.* 44. Up to this value the separation of the images may be of any desired amount. If the lines of greatest slope now correspond to the vertical plane the two images will be above one another, as shown in

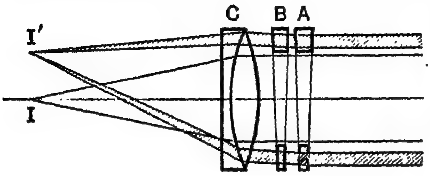


FIG. 44.

Fig. 45, and if the height of the smoke-stack above the water-line of the vessel sighted is known, a setting can be made and the range read with considerable accuracy from the scale attached to the prisms. If now the two prisms are rotated together through a right angle horizontal measurements may be made, as in

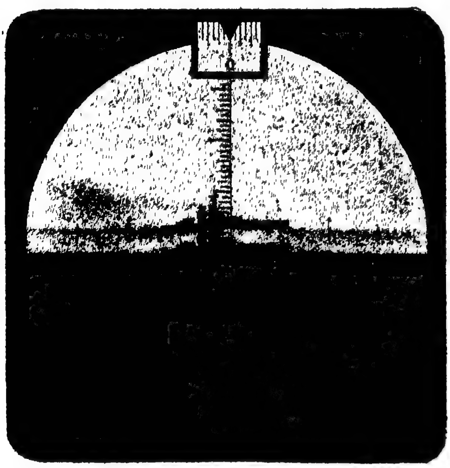


FIG. 45.

Fig. 46. As the apparent length of a vessel depends upon her course relative to the direction from her to the submarine, it is evident that when both the height and the length of the vessel are known the former is a suitable basis for the determination of her range, and the latter for finding her course when the range is known.

In addition to these particulars it is evidently of importance to find the speed of the vessel under observation. The time taken by the vessel to cross any fixed line in space is the time she takes to travel her own length, and thus if her length is known and the time taken to cross any fixed line is observed her speed can be determined. It is not satisfactory to take the projection of a line fixed relatively to the periscope as the line fixed in space for this purpose, as the errors caused by movements of the submarine and of the periscope itself may be considerable. Messrs. Barr & Stroud

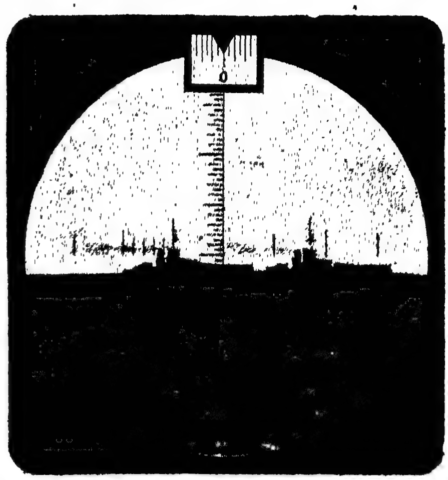


FIG. 46.

have overcome this difficulty by projecting into the field of view the image of a line which is controlled in direction by a small spring-driven gyrostat attached to the lower end of the periscope. *Fig.* 47 illustrates an instru-

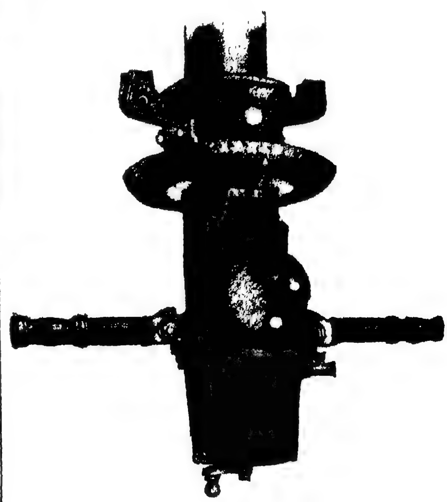


FIG. 47.

ment fitted with this device, the handle seen at the bottom serving to start the gyro. The system employed is shown diagrammatically in *Fig.* 48. A is the gyro pivoted on the bracket N, and B is the mark controlled by it,

the image of which serves for the line fixed in space. It is illuminated by the lamp C, and the light is introduced into the field of view by means of the prism D and the transparent reflector F. By means of the lens E, which must be arranged to give the proper magnifica-

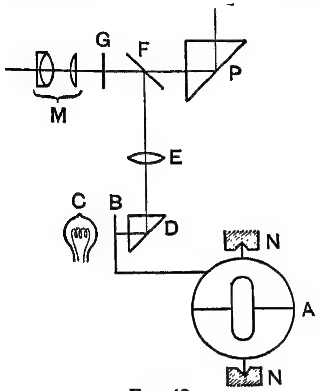


FIG. 48.

tion to the movement of the image of B as seen in the periscope, the mark appears focussed at G in coincidence with the images of external objects. The bracket carrying the gyro and the whole optical system are, of course, rigidly attached to the periscope.

When more precise range measurements than are possible on the principles utilised in the instruments just described are desired it be-

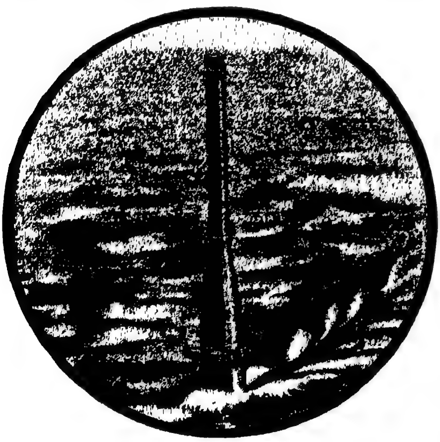


FIG. 49.

comes necessary to derive distances by observing the difference in direction from two stations on the submarine at a known distance from one another. Either the coincidence or the stereoscopic principle may be used to make settings as in ordinary range-finders (*q.v.*), and in the former case the base may be either horizontal, as with instruments used on land, or vertical, the arrangement which the shape of an ordinary periscope suggests as the more convenient. All these alternatives have been embodied in periscope range-finders. *Fig.* 49 shows the upper part of a vertical base range-finder, in which the two entrance windows, the separation between which is the base from which the measurements are made, may easily be recognised. The lower end is shown in *Fig.* 50, the handle on the right serving to effect coincidence, and the range scale being shown to the right of the eyepiece. The appearance of the field of view before

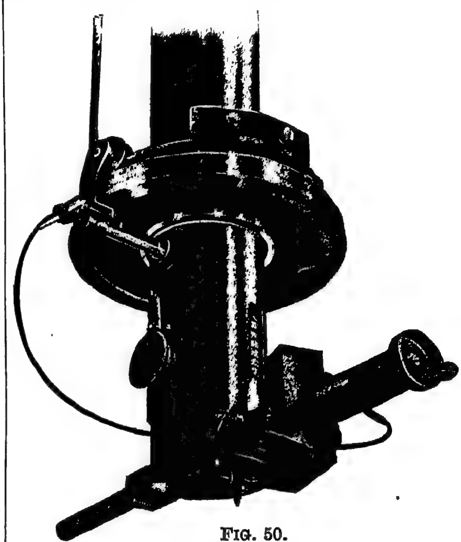


FIG. 50.

coincidence has been effected is shown in *Fig.* 51. The view seen through the lower window is brought to the central rectangular strip, and is surrounded by that from the upper window. An important advantage of the vertical base is that coincidence can be effected in spite of movements of the image due to the roll of the submarine, provided this is not so great as to cause the image to disappear entirely from the field of view.

The vertical base is not free from a number of objections, and to avoid these a horizontal base must be used. The difficulties caused by the rolling of the ship are overcome by special features in this range-finder which enable the image to be kept in the centre of the field of view. *Fig.* 52 shows the lower end of such a range-finder having a base of length 2·7 metres and a vertical length of 6 metres. This instrument is intended to be operated by two observers, and is fitted with a number of special devices for increasing the rapidity by which ranges can be transmitted. The magnification

is 20, and the accuracy under good conditions of seeing is 24 metres at 5000 metres, and 1

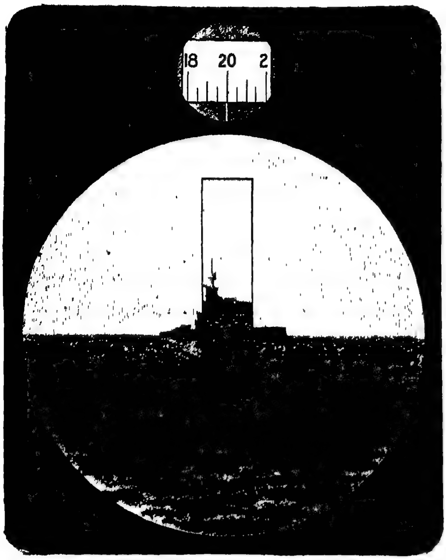


FIG. 51.

metre at 1000 metres. Of the two eyepieces seen, the right is that of the range-finder, that

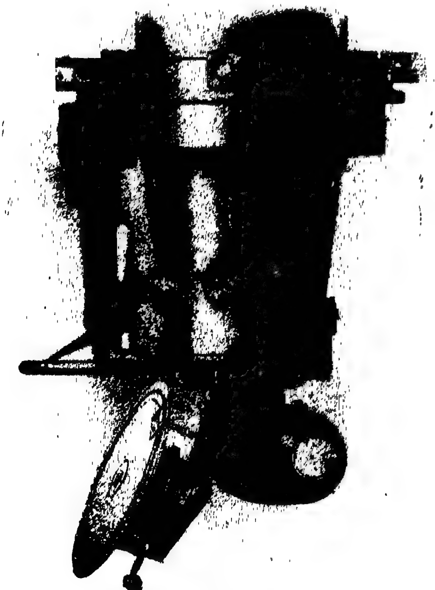


FIG. 52.

on the left being available when required for taking direct readings on the range scale. In addition there is on the other side a periscope eyepiece for a separate trainer. The main periscope tube thus contains two separate periscopic systems, one for the range-finder and one for the trainer periscope.

Fig. 53 illustrates another horizontal base instrument, and gives a general idea of the

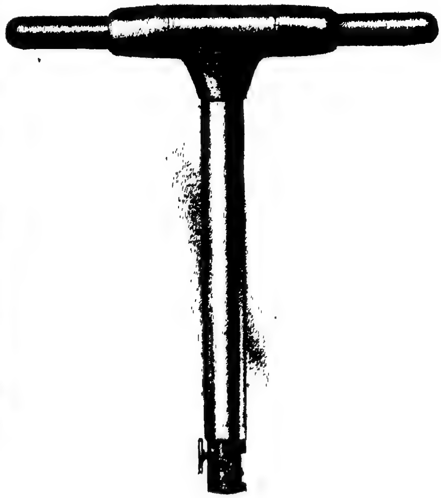


FIG. 53.

appearance of a complete instrument of the type. In this case the stereoscopic principle is employed, and the two eyepieces are clearly seen in the illustration.

T. S.

PETZVAL'S CONDITION FOR FLATNESS OF THE FIELD OF AN OPTICAL INSTRUMENT. See "Telescope," § (3); also "Optical Calculations," § (16).

PHANTOM RING. See "Navigation and Navigational Instruments," § (14).

PHONODEIK, MILLER'S: an apparatus for the demonstration and photography of sound curves. See "Sound," § (60).

PHONOGRAPH AND GRAMOPHONE, THE

§ (1) HISTORICAL.—The first mechanical reproduction of sound was effected by Edison in 1877. He wrapped a sheet of tinfoil round a cylinder and obtained a trace upon it of the movement of a diaphragm by a point which produced a scratch of varying depth as the cylinder was rotated. On allowing the point to travel a second time over the same groove the sound was reproduced, though, of course, very imperfectly. A horn was used in reproducing in order to intensify the

sound. In order that the tinfoil might be more easily indented by the recording point the cylinder had a groove under the foil where the trace was to be made. It was also mounted on a long screw so that the record was formed on a helix. Edison further suggested that the record might be made upon a flat disc and that both sides might be used.

In a later machine the recording style was short and was carried at the end of a spring, which was connected to the thin metallic membrane by two rubber buttons, which transmitted the vibrations of the membrane and at the same time damped the natural vibrations of the spring.

Edison himself dropped the phonograph as the result was so poor. It was Graham Bell, together with Chichester Bell and C. Tainter, who as the Volta Company went on with the idea. It was they who substituted solid wax for the tinfoil or waxed paper, and who cut away the wax with a sharp sapphire point instead of indenting it, and later used a chisel-end to produce a flat bottom to the groove. They, however, dissolved in 1885, and then Edison, going on from where they left off, made a successful machine and created the boom. The pitch of the screw was made much finer to increase the number of revolutions that could be made before reaching the end of the cylinder. Later the wax cylinders were made thin strengthened with ribs, the mandril on which they fitted was made taper, and they were driven by clockwork. Also the cylinders had no traversing movement, but the diaphragm was moved along by a screw as the cylinder rotated. The diaphragms were made of glass and had the stylus attached to them directly. Among other interesting devices, Edison, instead of black-leading the master wax cylinder, covered it with a film of gold by passing an electric current through a gold wire *in vacuo*, using this film to make the wax conducting.

The chief drawbacks of these instruments were the rapid deterioration of the records due to their soft material, and the amount of space required for their storage; moreover, the groove could only move the diaphragm in one direction—upwards—the reverse motion had to be produced by the spring of the diaphragm itself. Thus it had to press somewhat heavily upon the wax to enable it to follow the rapid vibrations of speech.

§ (2) THE PHONOGRAPH RECORDER.—These defects were largely overcome by making the record a transverse one, that is to say, one in which the groove, instead of varying in depth, is sinuous with approximately uniform width and depth; thus if a needle-point is placed in the groove, it will be moved to and fro sideways, the whole movement being imparted by the groove itself. The needle lies in a plane parallel to that of the diaphragm producing the sound—not perpendicular to it as in the phonograph; it is pivoted about its middle point, and while one end lies in the

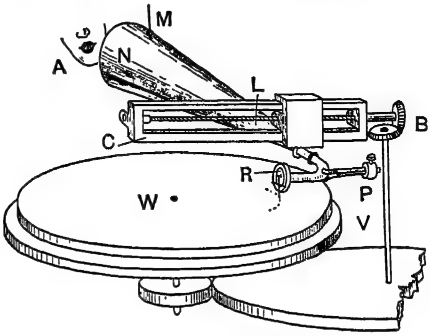


FIG. 1.—The Recorder.

groove the other is connected to the centre of the diaphragm (see *Fig.* 3).

When obtaining a record the singer (or instrument) faces the mouth of a large horn LN, *Fig.* 1, which is suspended by a string M at the mouth and is connected at the small end by a swivelling tube to the recording box R. This box is nearly counterpoised by a weight P. The box R with its connections is carried by an arm which can be traversed along a slide C by a screw S. The disc of wax upon which the record is to be taken is mounted on a turn-table V and rotated uniformly by a clockwork usually driven by a weight and connected through the bevel wheels B with the screw S. In this way the stylus cuts a spiral groove on the wax.

The stylus CD, *Fig.* 2, is brazed to the rod AB and strengthened by a rod EF. This little framework is pivoted at A and B, so as to rock about AB (which is parallel to the plane of the diaphragm) as axis. C is connected by a short wire to the centre of the diaphragm, which is of glass or mica, and to which it is cemented with wax; the end D is hardened and sharpened and cuts the trace on the wax disc. This disc is made of a mixture of waxes—ozokerite 2 parts and paraffin wax 1 part—or of a soap composed of stearic acid, caustic soda, and aluminium-hydrate mixed with ceresine or Japan wax; it must cut cleanly and not be too soft nor too brittle; it is warmed either as a whole in an oven or locally (by an electric current) while it is being cut. The wax disc is next black-leaded and an electrotype made from it. This electrotype (or a copy of it) is

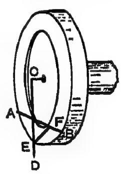


FIG. 2.

used to form the records. The records themselves are made of a composition of an inert powder and cotton fluff, with shellac or some similar material as a binder. The exact formulae used for the wax discs and the records are naturally kept secret by the manufacturers. After the composition for the records has been well mixed in the powdered form, it is rolled between hot rollers into sheets, cut into suitable-sized pieces, and pressed between the heated electros above described. The records so made are in the form of circular discs, usually either 10 or 12 inches in diameter, with a spiral trace having about four turns to the mm., commencing about 1 cm. from the edge and ending about 5 to 7 cm. from the centre.

§ (3) The Reproducer.—The reproducer or "speaker" is very similar in principle to the recorder. Instead of a sharp stylus, a rounded needlepoint slides in the groove on the record, with which it makes an angle of about 60°. This is hinged on a pair of V's against which it is held by two springs, so that it rotates through a very small angle about an axis, parallel to the plane of the diaphragm (*Fig.* 3); the other end is connected to the centre of the diaphragm by a small screw and a touch of wax. This diaphragm is usually of mica, it has a diameter of 4·5 to 6 cm., and is mounted between rubber rings called "gaskets" to prevent it from rattling, to damp its vibrations, and yet to allow it to move freely. Some diaphragms are made of glass, and it has been claimed that zirconium is a valuable ingredient in this glass. Others are of silk or other flexible material with a cork or card disc in the centre. The diaphragm forms the front of a short cylindrical box of magnalium or similar metal; the box has a metal tube in the face opposite to the diaphragm. This tube is inserted in the end of a pivoted conical tube called the "tone-arm," which terminates at its larger end in the beginning of the horn. The horn used to be a bell-shaped metal one, but it is now usually made of thin wood of rectangular section. It is found to give the best results if the passage from the sound-box to the horn has smooth bends everywhere. There is no mechanism required to cause the sound-box to follow the groove as in the phonograph, for if the tone-arm is freely suspended the groove itself is a sufficient guide. In order, however, that the needle may follow the groove easily it is obvious that it should always lie in a plane tangential to the groove. But as this latter is a spiral of varying diameter this can only be approximately achieved. Let P, Q, R be points on this spiral of which Q is on the middle ring, draw a tangent at Q, and with a point O on it as centre strike an arc of a circle, cutting the inner and outer rings of the spiral at P and R; it is obvious that OP and OR cannot be tangents at P and R. A better result is obtained by taking O on a line making about 14° with the tangent as at Q′ in *Fig.* 4. This circle will cut the spiral at points P′ and R′, at which tangents will make angles which are approximately the same with the lines OP′ and OR′; thus by setting the plane of the diaphragm at 14° with the line joining the centre of rotation of the tone-arm to the point of the needle, the latter will remain more nearly tangential to the groove throughout the reproduction.

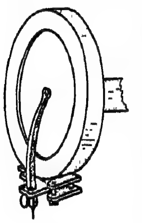


Fig. 3.

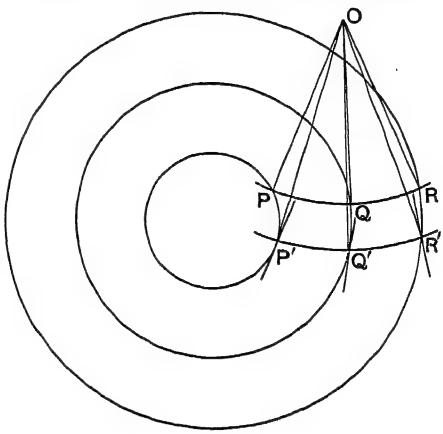


Fig. 4.

It is very important that the angular velocity of the cylinder, both in making the record and in reproducing it, shall be kept perfectly uniform, or the pitch will be affected. The linear velocity varies of course according to the diameter of the spiral, but as this variation of speed is the same in the reproduction as it was in the original recording, the resultant pitch is unaffected provided the angular velocity is constant. The linear velocity varies from about 120 cm. a second at the circumference to 40 cm. a second at the middle. By a recent invention the linear velocity is reduced to a uniform one of about 40 cm. a second—both in recording and reproducing the record—and thus the length of the record that can be made on a disc is greatly increased; the scratching noise is also said to be reduced by the diminution of the velocity.

Although the records are sufficiently good to afford a fair rendering of a musical composition or song, and so to give pleasure, they are far from perfect. The defects most commonly met with and the most difficult

to eliminate are (i.) a scratching noise, which is especially noticeable in the more delicate pieces; (ii.) an unpleasant and very obvious nasal tone, due no doubt to the horn; (iii.) a loss of many of the consonants in speech; (iv.) an inequality of rendering of the soloist and the accompaniment, due, of course, to the soloist being stationed immediately in front of the horn while the accompaniment is to one side.

In 1904 Parsons exhibited to the Royal Society a reproducer in which the needle, instead of being connected to a diaphragm, actuated a grid which covered a series of slots in a box, and thus opened and closed these slots. Compressed air was supplied to the box and issued from the slots into a horn when the slots were opened. Thus the record controlled the escape of air and produced in it variations which gave a powerful and yet good rendering of the record, comparatively free from scratch. From this has developed the Auxetophone and the Stentorphone.

The movement of the diaphragm of the recorder has also been used, generally with the aid of a mirror, to deflect a beam of light which was passing through an aperture on to a photographic film, and so to vary its intensity (Rankine increased the sensitiveness by projecting the image of one grid upon another grid); this photographic trace was afterwards used to control the light falling upon a selenium cell; the resultant varying current was passed through a telephone. A great number of attempts have been made to use some form of this method for the simultaneous recording of speech upon a cinematograph film, but so far without much success.

R. S. C.

PHOROMETRIC INSTRUMENTS: a type of subjective ophthalmic instruments used to deal with the measurement of ocular muscular want of balance and the muscular exercises for its correction. See "Ophthalmic Optical Apparatus," § (3).

PHOT: the unit of illumination on the C.G.S. system, equal to 1 lumen per square centimetre, or 10,000 lux. See "Photometry and Illumination," § (2).

PHOTO-CERAMIC METHODS OF GRATICULE PRODUCTION. See "Graticules."

PHOTO-ELECTRIC CELL: an instrument in which the Hallwachs effect is employed to measure light. See "Photometry and Illumination," § (34).

PHOTO-ELECTRIC EFFECT, application of quantum theory to. See "Quantum Theory," § (2).

PHOTO-ELECTRIC FATIGUE: the fall in sensitivity, with age, of a photo-electric cell. See "Photometry and Illumination," § (34).

PHOTO-ELECTRIC METHOD OF SPECTRO-PHOTOMETRY: the method depending on the use of photo-electric cells. See "Spectrophotometry," § (15).

PHOTOGRAPHIC APPARATUS

§ (1) INTRODUCTORY.—The ordinary process for the production of a monochrome picture of any object by means of photography consists of two operations—firstly, the making of a "negative," and secondly, that of the corresponding "positive."

The negative is a record of the form of the object expressed, as all monochrome pictures are expressed, in terms of gradations of light and shade, consequent on the exposure of a sensitive surface, composed of particles of a silver halide, to an image of the object in the camera, the changes induced in the sensitive substance being afterwards completed by the chemical process termed development. The positive or print in its simplest form is made by exposing to the action of light a similar sensitive surface under the negative, when there results, after development, a picture, the light and shade of which corresponds—within the limitations of the process—to that of the original object.

The sensitive substance now most generally employed for the purpose of making photographic records in the camera is silver bromide or bromo-iodide emulsified in gelatine. This emulsion is coated upon glass or celluloid and dried, and the coated plates or films come into commerce as "photographic dry plates" or "films" as the case may be. It is proposed in this section to consider the instruments, other than lenses,[1] by means of which photographs are made, and, in the main, those which are used for the production of negatives.

A photographic camera in its simplest form is a light tight, rectangular box terminating in two flat ends which are parallel. In the centre of one end there is pierced a very small aperture, the pinhole, which is then covered by a simple shield that is readily removable. If there now be placed inside the box (under suitable "safe" illumination), at the end opposite the aperture, a sensitive plate, all that is necessary for the production of a photograph has been provided. If the box be placed upon a rigid support in front of a well-illuminated object, on uncovering the aperture an image will be formed upon the plate. If exposure to this image be made for a sufficient length of time—if the conditions are suitable

[1] See "Photographic Lenses."

the period will not be unduly long—the plate will yield upon development a negative image of the object.

It is assumed often that satisfactory photographs cannot be made by the use of a "pinhole" in place of a lens, but this is not correct, and well-constructed pinholes are sometimes used on modern cameras for subjects where brief exposures are not necessary. Provided that the aperture be properly made and its diameter in relation to the plate distance fulfils the requirements of the optical theory governing the formation of pinhole images, little objection can be taken, so far as "definition" of the image is concerned, to photographs of still life, architecture, and landscapes, when meant for pictorial purposes, more especially if the photographs are fairly large in size. The interest of the "pinhole" is, however, mainly theoretical, and in practical everyday photography of different kinds, lenses are used for the formation of images in the camera.

It would be foreign to the purpose of this section to discuss the apparatus specially designed for the purpose of employing light sensitive materials as recorders in investigations in different branches of science and technology, for the apparatus varies greatly according to the nature of the inquiry which is to be made, and is generally peculiar to the particular research. Much laboratory apparatus is of standard type and is applicable to wide fields of endeavour, as for example the camera in conjunction with the telescope, the spectroscope, and the microscope, and in these respects other sections[1] should be consulted.

In many instances photographic apparatus is employed to record the behaviour of photographic material itself, as when the spectrum camera is used to ascertain the distribution of spectral sensitiveness in a dry plate or other sensitive surface, or the camera shows the appearance under the microscope of the silver bromide grains in a gelatino-bromide emulsion, and thus provides a permanent picture of certain aspects of its physical state.

As well as being an aid to a large and increasing degree in a number of branches of scientific and technical research, it should not be overlooked that photography of itself is an important industry—either alone or in conjunction with other crafts—utilising some of the finest products of optical, chemical, and mechanical skill. With its growth there has come specialisation, and each branch has evolved forms of construction to meet its own needs. No one camera will meet conveniently all the requirements in the many activities of modern photography. In view of the specialisation ordained by these requirements, it seems desirable that the essentials for an ordinary camera in its simple form should be stated. Cameras are essentially chambers for the purpose of exposing sensitive plates to the action of luminous images formed by optical means. Disregarding for the moment the precise mechanical form of the construction, a camera of simple type as used in ordinary photography consists of a support for the lens (the "front"), used to form the image of the object; another support (the "back") for the movable receptacle (the dark slide), usually employed for holding the sensitive plate, and a device (the "bellows") for preventing any light reaching the plate other than that which forms the image, the distance between the "front" and the "back" being adjustable, while parallelism is maintained. To enable the image to be examined and adjustment made, a screen of ground glass ("focussing screen") is arranged in a movable frame in the back. To permit these adjustments to be made, certain devices for movements in the "back" (reversing frame and swing-backs) and in the "front" ("rising and falling" and sometimes sliding and cross-fronts) are provided. During exposure the sensitive plate in its receptacle occupies the position previously held by the screen and therefore receives a similar image. The camera is supported upon a stand, by means of which its position relative to the object is adjusted and maintained.

The most important items of photographic apparatus, excluding lenses[2] and correction and colour filters,[3] are

Cameras,
Finders, plate-holders, and changers,
Exposing shutters,
Stands,
Enlargers,

and in every division there are differences in type and a great variety of patterns following the types.

§ (2) Cameras.—Discarding arbitrary distinctions as far as possible, and for the purpose of description, cameras may be classified as follows:

Stand Cameras.
- Field or portable.
- Studio.
- Copying (including cameras specially designed for photo-mechanical photography).
- Kinematograph.

Hand Cameras.
- Plate or "cut film."
- Rollable film.

Aerial Cameras.

[1] See "Telescope," "Microscope, Optics of the."

[2] See "Photographic Lenses."
[3] See "Light Filters."

This differentiation should not be taken to mean that in no case can a camera designed specially for one purpose be used for another. In some cases there are limitations, but not necessarily exclusions. A field camera, for example, could be used in the portrait studio, but it would not be so suitable for many reasons as the studio type. Even though the design of the field camera were good it lacks "stiffness" (due mainly to reduced material) and, in consequence, would be strained by the constant stress of the lens suitable for portraiture in the studio, which, being of considerable focal length and large aperture, is relatively heavy to the lenses used normally on a camera of the field type. The swing-backs on the field camera would not be so convenient for rapid manipulation as those on a good studio camera, and the field camera lacks certain desirable accessories as, for example, the repeating back. The studio camera of high grade is a heavy rigid instrument of considerable bulk, necessitating a strong stand of a different pattern to the portable camera—and it cannot be folded for transport. It would in consequence be extremely inconvenient "in the field."

(i.) *Field or Portable Cameras.*—Two cameras of the original patterns of this class are shown in *Figs.* 1-4. *Figs.* 1-2 represent the "square bellows, heavy-form," and *Figs.* 3-4 the

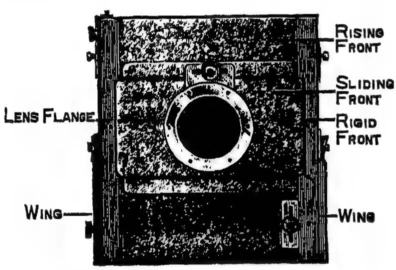


Fig. 1.

"conical bellows, light-form," instrument. The "heavy-form" camera of the best makers is still practically the same as when introduced,

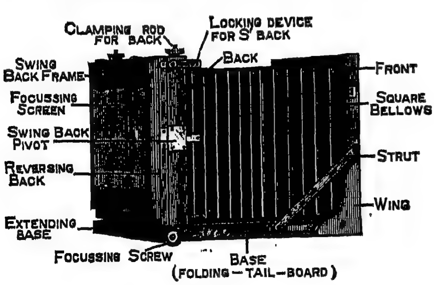


Fig. 2.

but the "light form" has undergone several modifications with the object of reducing weight and bulk. The cameras following these forms differ mainly with respect to their focussing arrangement, which is the governing factor. In the "heavy form" the back is movable and the front is fixed, the side of the

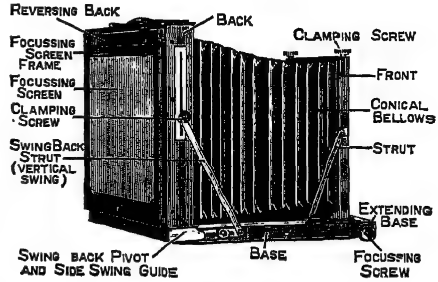


Fig. 3.

bellows being parallel, whilst in the "light form" the front is movable and the back stationary, the bellows taking the form of a

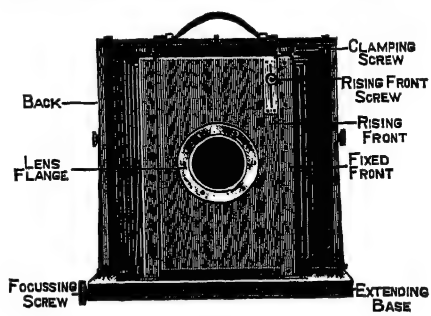


Fig. 4.

rectangular cone, and this construction permits greater compactness when the camera is folded.

The front and back of the camera should move to and fro parallel to each other, and this is provided for in the construction, the movement being generally effected by means of a rack and pinion. The "extension" should preferably be not less than twice the length of the diagonal of the largest size of plate that the camera is made to take. In many cameras greater extensions are possible, but the cameras used in that state under field conditions are never so rigid as when more closed, and in some "light-form" instruments are unreliable even before the limit of the extension is reached. The normal position for the lens is opposite to the centre of the plate; the prolongation of its axis should meet the plate centre and be perpendicular to its surface. In practice variations are necessary. To give the required movement in a vertical direction a panel (rising and falling front) fits into a groove in the fixed front, and to this the lens is attached, and by this device

movement up and down is readily effected. The rise should be greater than the fall. In the camera shown in *Fig.* 1, $8\frac{1}{2} \times 6\frac{1}{2}$, the rise is 2 in. and the fall 1 in., and in *Fig.* 4, 12×10, 3 in. and $1\frac{1}{2}$ in. respectively. To secure a lateral movement (sliding or cross-front) another panel is fitted to the vertical front, and to this the lens is attached. The rising and falling front is a necessary adjunct, but the sliding front is much less important and may be eliminated without serious inconvenience. Movements of the lens made by means of the cross-fronts are equivalent in effect to lateral movements of the plate in the opposite directions. To effect movements of the plate would mean an impracticable complication. The front in a "light-form" camera is usually brass strutted to secure rigidity, a quite necessary provision in the larger sizes. Where more than one lens is used the front should be pierced to take the flange of the larger, and the smaller is fitted when required by means of a brass adapting ring.

When the front of the camera is set back to permit the use of lenses of short focal distance the conical bellows tends to crowd to the back and so cut off image from the margins of the plate. This drawback is prevented by providing "tapes" on the sides of the bellows, by means of which external connection is made to the front, and the bellows are thus drawn forward. Photographic dry-plates are rectangular and are made to standard sizes, and with one exception (lantern slide plates, British $3\frac{1}{4} \times 3\frac{1}{4}$ in.) have dimensions longer on one of their sides than on the other. To meet this condition cameras are made with the backs square and are fitted with a removable frame carrying an inner (generally hinged) frame bearing a focussing screen equal in size to the largest size plate that the camera is made to take. The outer frame or reversing back being square and removable, it may be placed in the back with either dimension vertical or horizontal at will.

The receptacle (dark slide) for the sensitive plate replaces the focussing screen frame, which is turned out of the way and generally folds over the slide when in position.

Another movement of importance is the swing-back system, by means of which the plate can be tilted in a horizontal or a vertical direction. The swings in the camera shown in *Fig.* 3 take place from the hinges which attach the back to the base. In *Fig.* 2 the back is in two parts. The rear portion—to which the bellows is attached—draws out and swings from the centre on attachments to the fixed part. The vertical swing-back is used to correct distortion of perpendiculars due to camera tilt. Both swings are used, where necessary, as an aid in focussing.

Several changes have taken place in the early "light-form" camera. The solid baseboard has been reduced as much as practicable, the central portion being cut away, and in its place there is now fitted a light combined tripod head and turn-table. This eliminates the loose head which usually accompanies the tripod. The front is smaller and lighter and is arranged so that it can be swung. The rising panel remains, but the whole front (which carries the bellows) can be raised. This means that a greater rise can be obtained without interference by the bellows. A distinct improvement resulted from the addition of back-extension focussing. When the image is small in proportion to the object it is immaterial for all ordinary purposes whether the front or the back moves for focussing, but where the image is large, as when objects near to the lens are to be photographed, more particularly when the image is to be made to a predetermined size, front focussing is impracticable, and such cameras can only be used by a device which practically makes them back-focussing instruments. The duplex system is therefore a great advantage, and it moreover permits the use of lenses of short focal length to obtain pictures embracing wide angles.

The result of these changes has been to reduce considerably the weight and bulk of the camera, but it is proper to observe that such instruments are only reliable when their design is united with the best possible workmanship.

Cameras of the "heavy" type in modern use are exemplified by the "Premier" (Watson) and the "Square Bellows" (Ross), and of the "light" type by the "Acme"

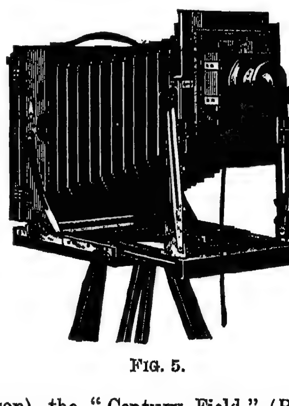


FIG. 5.

(Watson), the "Century Field" (Ross), and the "Sanderson," *Fig.* 5 (Houghton).

No departure of moment beyond those indicated has been made in portable-stand cameras during recent years, and it is probable that the patterns will persist. The high pitch of development reached

by the best forms of modern hand camera has caused these instruments to satisfy the requirements of some of the most exacting workers. The tendency since the introduction of the high-class anastigmat lens has been to make negatives of small size, and to enlarge the pictures when required, a tendency only accentuated as improvements in that instrument and in other directions take place. Consequently, it is not likely, in view of its convenience and excellence, that the small hand camera will be laid aside in favour of the instrument generally used in earlier days. In the more restricted fields of industrial photography, with its generally less adventurous task, as well as in many fields of technical work, the ordinary portable camera will continue to be used.

(ii.) *Studio Cameras.*—The cameras of standard form are heavily built, back focussing and non-folding. The instrument has long extension, fixed or movable front, which is provided with lens panel for slight rise and fall. In some of the better forms the front is movable by rack and pinion, so that front adjustment without moving the camera can be made, which is an advantage when copying. The lens is usually fitted to an inner "lens board," readily removable, behind which may be placed the exposing shutter, and this shutter being invisible enables exposing to be done without the knowledge of the sitter for a portrait. The arrangement of the back is different to that in a field camera. The size of camera selected for the portrait studio generally permits a fair range of sizes, say from 15/12, 12/10, or 10/8 down to ½-plate. In a particular studio a certain size will be more used than others, and will be generally smaller than the full size that the camera is intended to take. As it is the common practice in good portrait studios to take several pictures, and rapid manipulation is desirable, provision is made for this by means of an auxiliary holder ("repeating-back" type). The focussing screen for the full-size plate—usually a loose frame—is removed, and in its place is introduced a larger frame bearing a second focussing screen of the smaller size which slides between two parallel bars. The plate-holder ("dark slide"), containing space for two plates side by side, also fits these bars. The focussing screen is masked from the front by a wooden frame flush with the face of the repeating back (and changeable at will), so that the operator sees on the screen only an area corresponding to the plate that will be used. This plan is much more satisfactory than the ruled lines on the screen showing limits (which are sometimes used), and prevents confusion. When the image is focussed the screen is moved laterally (though it is sometimes hinged and is swung out of position) and the plate-holder takes its place, being "set" so that the first plate is in position. After exposure the plate-holder is moved along until the second plate is brought into position, and exposure is then made. The shutter or the plate-holder is fully drawn all the time, but no extraneous light reaches the plate, the repeating back acting as a shield. When the camera is one made to take large plates the repeater form of back may be employed, but this means considerable bulk. The repeating-back principle is frequently applied to cameras not intended for portraiture and is capable of considerable application. The plate-holders are made square (save small holders for repeaters) and plates are placed with the major dimension horizontal or vertical at will.

The back of the camera swings in both a horizontal and a vertical plane. These movements are important, being often required when focussing images of objects in relief near to the camera when lenses with large apertures are used, though their use in this respect is often abused. The support (stand) is frequently combined with the camera, which makes for convenience and rigidity. Provision is made for raising and lowering, by rack and pinion, and for tilting the camera out of the horizontal. Castors are provided to the feet of the stand for ease of movement. Otherwise, independent stands of similar construction are used. The combined apparatus is well shown in the "Improved Studio" (Ross), "Soho" (Marion), and "Combination Studio," *Fig.* 6 (Watson).

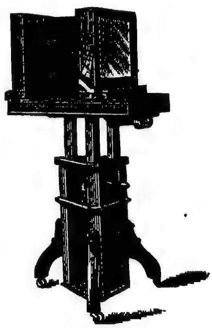


FIG. 6.

For special work, as in the portraiture of children, the reflex principle has been applied, as in the "Studio Minex" (Adams), and twin-lens cameras have also been used. These cameras have the advantage that the image can be viewed on a full-size screen, focussing effected, and exposure made at the precise moment when the appearance is judged to be satisfactory.

§ (3) COPYING CAMERAS.—The term "copying" is applied to that branch of work which deals with the photography of flat surfaces. For much copying the ordinary studio camera is quite satisfactory.

In pure photography the most important branch of copying is that of making transcripts of pictures of different kinds. For picture copying in galleries, museums, and private collections, the "heavy-form" portable camera, provided that the instrument is an efficient one of its kind and a suitable rigid

stand is employed, meets the requirements. Cameras of heavier construction and greater refinement, if combined with special stands, would confer certain technical advantages, but the inconvenience and cost of transport would mean serious objections. In some continental photographic houses where picture copying is a feature, apparatus of a special character is sometimes employed where the camera is the dark room. This camera-dark room (in which the operator works) and the picture-supporting easel are built on a platform mounted upon a "railway" turn-table, which permits the whole system to be rotated to allow for changes in the sun's position.

The greater part of copying work comes in the domain of photo-mechanical photography,[1] which is the branch of the craft producing reproductions or transcripts in various forms of different kinds of picture—using the term in a broad sense—by distinct methods differing from those of pure photography, which are finally associated with some form of the printing press (photogravure, photo-lithography, collotype, line and half-tone process block making in monochrome and colour and combinations thereof), and for this certain forms of copying camera have been devised. Great accuracy is required in the copying of maps and plans and of pictures for colour block making and photo-lithography, and for this work the modern camera has become to some extent an engineer's construction.

(i.) *Details of the Camera.*—A camera system for the purpose of "copying" should fulfil the essentials of parallelism between the "front" supporting the lens and the surface of the copy board bearing the subject to be copied (the "original"), and between the lens "front" and the focussing screen. If by proper construction these conditions are satisfied, the principal axis of the lens will be perpendicular to the surface of the plate and to the surface of the "original." If, however, negatives laterally reversed are to be made (which is necessary in many processes), the focussing screen must be at right angles to the copy board, and a right-angle reflecting prism, silvered on its hypotenuse, must be mounted in front of the lens, the reflecting surface being at an angle of 45° to the lens axis. In place of the prism, a plane mirror, silvered on its surface, is frequently used.

The camera and the copy board must be supported so that the centres of the focussing screen and the copy board, respectively, are in line with the lens axis. When this is done the image of an "original" which is placed on the centre of the copy board will fall in the centre of the focussing screen. If the camera be then rotated horizontally 90° (the prism being mounted in front of the lens) and the board be shifted laterally in a plane at right angles to the focussing screen to a sufficient extent, the image of the "original" will again be centred on the screen, but it will now be laterally reversed. Further, the camera and copy board must be so arranged in relation that, at desire, either of these conditions can be maintained, whatever be the distance between the lens and the "original," *i.e.* as the camera is moved to and fro by the operator for various sizes of image. By arrangements conforming to the conditions stated, "copying" can be done with accuracy and with the rapidity necessary to meet the economic and other conditions which obtain. A suitable method for large copying systems, say, for plates 36 × 24 in. and below, is to mount the camera and the copy board upon rigid stands, the feet of these being fitted with grooved wheels which engage on parallel metal rails laid on the floor of the studio, or a pair of rails may be laid at right angles to one of the walls of the studio to take the camera stand whilst the copy board is fixed upon the wall, provision being made for its movement laterally for centring when the camera is turned into the position (previously described) for making reversed negatives. In some studios where very large drawings have regularly to be copied a large area of the wall is prepared as a copy board, and there is then sufficient room to accommodate drawings when the camera is arranged for either "direct" or reversed negative making. Great care is necessary in the construction of boards to ensure flat surfaces which do not warp—a problem of difficulty, more especially when they are exposed to the heat of powerful arc lamps as in artificial light studios. The floor of the studio should be laid firmly upon a suitable foundation to avoid vibration. The camera stands and copy-board stands are made of wood (preferably teak) and metal, but the most recent practice is to use only metal. Cameras constructed for map and plan copying (which are suitable also for general copying) of large size are shown in *Fig.* 7, "Calcutta" pattern (Penrose), and the "Precision" pattern (Penrose). The stand in *Fig.* 7 is of wood and metal; in another pattern a metal stand is employed. Both were constructed for the Survey of India office and are in use in many similar establishments. In some studios when large negatives are made the "dark room" is the camera. The lens is mounted on a suitable panel fixed in the wall of the room. The "original" to be photographed is supported on a copy board fixed upon an easel which travels on rails on the floor outside the room, whilst the focussing screen (and eventually the sensitive plate)

[1] See "Photography and the Printing Press," *J. Roy. Scottish Soc. Arts*, "Keith" Lectures, Nos. 10, 11, xvii.; *Dict. App. Chem.*, Thorpe.

rests in a holder mounted upon a stand having a similar traverse on the floor of the dark room. The holder can be moved upon the

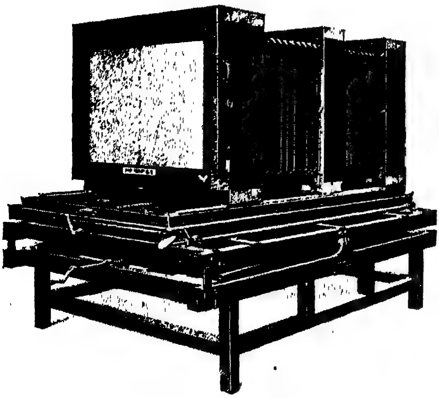


FIG. 7.

stand by means of a screw, and is itself fitted with adjusting screws for producing parallelism with the copy board. Setting for size of image is effected by moving the stand, the final focussing being made by the screw upon the stand which moves the plate holder.

A brief allusion has already been made (*q.v.*) to photo-mechanical printing processes and the difference between these and processes in pure or ordinary photography. For a more detailed explanation the reference given should be consulted. Maps, plans, and drawings in line (broken tone) of different kinds, drawings in colour, and oil paintings, nature photographs, where the light and shade is continuous (continuous tone), may all be copied in the same apparatus (the ordinary camera copying system as described), provided that the positive or print method for which the negative is to be used is confined to processes of pure photography [1] and to certain others in photo-mechanical photography, but not to all. Briefly, the visual effect of the continuous shading or continuous tone in a wash drawing or a nature photograph, for example, can only be produced by certain important methods by translating these shadings into "broken" tone. This translation may be, and generally is, effected by optical means by the application of the half-tone principle of Ives. The most important process where this principle is used is in the ordinary type-high half-tone process block which supplies the great majority of illustrations accompanying printed matter of various kinds. The translation of tone is produced in the copy negative, and is brought about by placing in front of the sensitive plate during exposure the so-called "ruled screen." The ordinary ruled screen consists of two rectangular plates of transparent colourless glass—cemented together in optical contact—bearing upon their inner surface alternate clear and opaque lines usually of equal width. The lines form angles of 45° with the sides of the plates, but in each one of the pair the orientation of the lines is the opposite of that in the other. In consequence, when the plates are in facial contact there is formed a series of uniformly distributed transparent spaces, rectangular, and of equal size.

The effect of the screen is to break up the continuous tones of the original into dots of various sizes, and in the completed printed picture the illusion of shading is produced by optical mixture.

The surface of the screen when in use must be parallel to the surface of the sensitive plate, the interval between the two being of small order. This interval is, however, not a constant amount whenever a "copy" negative is made, but is varied as required in harmony with other conditions, the screen distance being only one factor out of the total number to be considered. The screen plate must be held rigidly in a smooth-working traversing frame fixed on a support inside the camera, which frame can be caused to move to and fro at right angles to the lens axis and accurately set within the limits required. The traversing frame holding the screen is moved on its support for adjustment as to screen "distance" by lever and micrometer screw, the screw providing an adjustable limit to the position of the lever. This limit is the correct screen distance or "setting" as determined by examination of the image on the focussing screen, or by the particular system of working adopted by the operator, and this distance can be read off from the indicator scale provided. If the screen distance has once been set the lever can be moved to throw the screen back out of position, but on reversing the direction of the lever the screen can be brought back with exactitude to the original "setting," because the position of the micrometer screw provides the limit to its movement. A movement of this kind is necessary, because when the screen has been set by examination it is necessary to throw it back for the insertion of the plate holder (dark slide), and when the shutter of this slide is withdrawn to prepare for exposure of the plate the screen must be brought forward to the predetermined position in front of the sensitive plate, and thrown out of position when the slide is to be removed from the camera. If, on the other hand, the screen is "set" to a distance by any other system than visual examination of the image, it suffices to fix the micrometer screw to the required dis-

[1] See "Photography, Photo-Mechanical Processes," *Dict. App. Chem.*, Thorpe, 2nd ed. iv. 230.

tance by the indicator, and the screen is brought up to the required position when the shutter of the plate holder *in situ* in the camera has been withdrawn. When not in use the screen holder is thrown back "out of action," and if the camera is to be used for ordinary copying without the screen the latter is removed from its holder.

Cameras with this provision are termed "process" cameras. So far as the camera *qua* camera is concerned the presence of the screen gear is not a detriment to its use for other copying purposes, which gain indeed from the accuracy and rigidity and convenience in movements given to the instrument to fit it for its special purpose. During recent years very great advances have taken place in this branch of camera instruction, largely because it was recognised from the first that the problems to be solved were essentially those for the engineer. As far as practicable, wood as an element in the construction of cameras and plate holders has been replaced by metal. To obtain the necessary accuracy in cameras of this type and to maintain it, strength combined with good design and workmanship is required. In many studios, particularly those situated in hot and moist climates, the conditions are extremely trying to apparatus and personnel, and these factors are recognised. Modern constructions are made as far as practicable to meet the most difficult conditions, and the adoption of a standardised pattern designed to meet not the normal but the abnormal is economically sound from every point of view, when, as in this instance, it means, principally, good design and an intelligent choice of material.

In modern process cameras of the best class the back and front are separately mounted on iron base plates which travel upon steel runners, the distance apart of the two being adjusted by means of screws, and a screw and bevel gear is employed for adjusting the lens panel. These screws are all worked from the back of the camera by means of one removable handle. The focussing screen and lens board are detachable. The plate holder or dark slide is directly interchangeable with the focussing-screen frame. The plate holders contain adjusting bars for holding different sizes of sensitive plate. When, however, a large camera is not employed for the full or the larger sizes of plate for which it is constructed, an adapting frame ("Studio Cameras," *q.v.*) is fitted in place of the full-sized focussing screen. The adapter takes a smaller focussing screen and a corresponding plate holder. This practice considerably reduces the fatigue in manipulating large and heavy dark slides.

The camera stand and easel for the copy board are generally in one construction. The precise form varies, but essentially the arrangement consists of a pair of parallel rails framed up in a cradle, which is suspended by springs from a second and rigid construction resting upon the floor. The easel is fixed at one end of the sprung cradle and the camera is supported upon a travelling carriage, which is provided with a turn-table to enable the camera to be rotated into the correct position for reversed negative making. The camera in its carriage will travel to and fro easily (maintaining correct relation with the copy board), and the arrangement may be clamped firmly in any position. The copy board is of wood, articulated, to avoid as far as practicable the effect of warping, any shrinking occurring being taken up by means of the screws in its framing. It runs on the guides which form part of the easel, and the board is fitted with devices for securing parallelism with the focussing screen, and, since arc lamps are usually used in reproduction studios for illuminating originals to be photographed, it has been found convenient to attach adjustable supports for these lamps to the camera stand.

The best modern stands are wholly of metal, the only wood employed being in the copy board. The "springing" of the cradle bearing camera and copy board is for the purpose of avoiding transfer of vibration from floor to camera system, but even the best arrangements do not wholly avoid this trouble under certain circumstances, and great care is required, especially when the exposures are prolonged.

The "Empire" process camera, *Fig.* 8, and

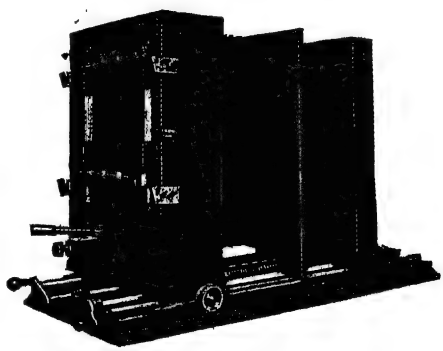


FIG. 8.

"Empire All Metal Stand" (Penrose) (not shown), and the "Arc Gear" process camera (Hunter), *Fig.* 9, are good modern examples of these constructions.

(ii.) *Automatic Focussing.*—For rapidity of work it is desirable that some form of automatic setting for the respective focal conjugates be employed in copying systems, and attention has been given during recent years

to the best means of using the principle. Given the focal length of the lens accurately, and the position of its nodes being known,

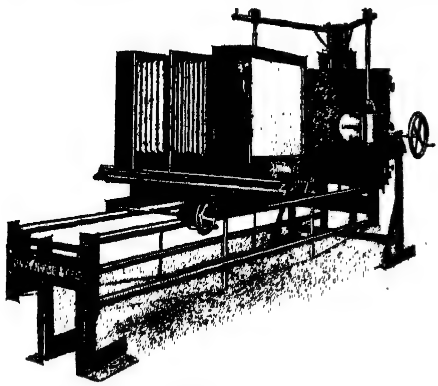


Fig. 9.

scales can be prepared and adjusted, so that by setting the respective index pointers, adjusting for size and sharp image, it would appear, could readily be done. In practice, however, the employment of this plan is not so simple. It is, however, practicable to use scale setting as a close approximation and finally focussing in the ordinary way. The scales for any apparatus are conveniently prepared by trial. Such a system can be made to effect a great saving of time. More satisfactory is a linkage system by means of which, on focussing, the image varies in size, but is always "sharp." This plan is adopted in certain forms of camera, but it entails considerable cost.

(iii.) *Reversing Devices.*—Reference has been made to the production of laterally reversed negatives which are necessary in certain photographic processes. In the wet collodion and collodion emulsion processes the negative may be made direct and reversed by stripping the film from the support, turning it over, and causing it to adhere to the glass, or by suitable procedure it may be converted into an unsupported flexible film, and the same plan has been largely adopted with the so-called stripping dry plates. Reversal, in practice, may be also effected by the use of certain celluloid films on which negatives are produced thin enough to be printed from either side. Another method especially convenient with dry plates when used in portable cameras employed away from the studio is to reverse the plate in the dark slide so that the glass side instead of film side is towards the lens, allowance being made when focussing. But the general practice is to use the optical devices referred to previously, viz. the right angle prism or the plane reflecting surface.

There have been at all times difficulties in obtaining as a regular supply, blocks of glass sufficiently free from internal strain and striae to make prisms of more than 3 in. face which could be used without detriment in conjunction with high-grade lenses, the production of blocks for larger sizes being often a matter of uncertainty and considerable time. Prisms have, however, been made of 5½, 6½, and even 8 in. faces. The mounting of the prism of moderate size is not difficult, with large sizes by no means easy. Prisms are mounted in triangular metal boxes with an external ring bearing a screw thread which engages into the lens mount, the face of prism and front combination being as close as practicable. The weight of the prism is borne by the lens mount, a system to which objection might be taken. To adjust accurately the face of the prism parallel to the copy board, probably the best way is to attach the lens to the front by means of a flange supplied with a rotating collar provided with a clamp (Zeiss). The same result can be brought about and the strain of the prism weight can be taken off the lens by means of an adjustable bracket, on which the prism mount rests, fixed to the lens board. In this way horizontal tilt of the prism can be entirely prevented, and there is no tendency to "sag."

Mirrors are plain glass or metal. The glass is surface silvered. The surface of the silver tarnishes and requires re-polishing, and being delicate is easily damaged. Metal mirrors are more durable if care be taken, but they suffer in acid atmospheres, and especially when the air is humid, but are costly, and those produced to the time of writing (1922) have not been perfectly satisfactory. Efforts are being made to employ chromium steel, and the results are promising. Platinised glass mirrors are at present the subject of experiment, but the difficulties connected with the vitrification have not been overcome. Mirrors are mounted in right-angled boxes of wood, of wood and metal, and of metal alone. Boxes of wood, which are largely used, are open to considerable objection. The best form of box is of metal with cast frame, and provided with adjusting screws to set the mirror with respect to the lens. The mirror is mounted behind the lens. This, although not absolutely essential, is necessary in practice. The mirror is in this way protected from extraneous light—a very important condition—and from the possibility of mechanical damage.

A serious objection to the prism arises from the absorption of photographically active light, which increases rapidly as the mass of the glass increases. This applies more especially when the process is "wet collodion" and the source of illumination is enclosed arc lamps, owing to the particular spectral sensitive-

ness of the silver salt. This cannot be avoided, but it may be minimised by a careful selection of the glass. Loss also occurs with metallic mirrors, especially when tarnished. With glass mirrors silvered on the surface the loss is appreciable even when the tarnish is very slight.

(iv.) *Colour Screens.* — Provision must be made for the support of colour screens, either the correction filter—" yellow " screen—employed in conjunction with orthochromatic or panchromatic dry plates, or for the selection filters [1] employed in the three-colour process, or any method where the colour luminosities of the subject are to be controlled.

In ordinary photography with portable cameras simple screen holders provided with caps may be attached to the lens, or the filter may be mounted behind the lens board by any convenient device, the necessities being freedom from stress on the filter and that it be supported with filter face at right angles to the lens axis. In the case of reproduction cameras where filters are made with optically worked glass " flats," great care is necessary, and it is now usual to employ a filter holder as part of the lens front, the carrier for the filter flat sliding behind the lens.

(v.) *Ruled Screen Carriers for Special Purposes.*—It frequently happens that a picture made by a half-tone (ruled screen) process is to be copied in the camera for the making of another printing surface by the same principle. Thus, for example, the only available picture of which a " half-tone " block is desired may be itself an impression from a " half-tone " block. A translation made in this way will always produce a " pattern," [2] which " pattern " may in certain conditions amount to a pronounced moiré. This phenomenon varies in appearance as the angle between the existing screen lines in the print, and those in the screen used when making the second negative, alters. The " pattern " can never be in practice eliminated, but the objectionable effect may be minimised to an extent that it becomes inoffensive by adjustment of the screen when making the copy negative.

Similarly, when the screen block process is employed for producing transcripts of pictures in colour, as in the three-colour block or other screen process methods, impressions from different surfaces in their respective colours are superimposed. Consequently, to avoid the moiré effect and produce a visually negligible pattern, the angles of crossing of the screen lines of the components must be adjusted. The most convenient plan is to employ a circular screen fitted in a rotating carrier mounted in the screen gear frame, when the suitable angular differences for the separate negatives can be made.

§ (4) Self-contained Camera Copying Systems.—Photography is widely employed to-day for the purpose of rapidly multiplying copies of letters, reports and documents of all kinds, printed matter, drawings, and matter of a like nature. Hand copying in many instances would be slow and expensive, and would entail always careful checking, whereas the photographic copy is in essentials free from error. To reduce operations to a minimum, and to enable the work to be done rapidly and without photographic skill, the apparatus—of which the " Photostat " (Alfred Herbert, Ltd.) is the most important example —is made complete in itself, and the operations of arranging the original, lighting, focussing to scale, exposure, development, and finishing are practically automatic. The copies are rapidly made upon sensitive coated paper which is in continuous roll form in the apparatus, negative pictures resulting. In the greater number of cases a negative copy answers the purpose, but if positives be necessary the negative is copied, when a positive of the original subject results.

§ (5) Plate Holders and Changers. (i.) *Holders.*—The plate holder or dark slide usually employed with portable field cameras is that known as the " double book form." This holds two plates back to back, separated by a diaphragm, and opens like a book. Movable shutters draw out to expose the plate when in the camera, which shutters fold back over the slide.

Solid (non-opening) slides for two plates are now largely in use, particularly for small sizes, and when well made are satisfactory, possessing many advantages—lessened weight and bulk, and greater protection of the plate from light and from the possibility of accidental exposure due to careless handling. The slides are filled from the front by withdrawing the exposing shutters, and, in this respect, with many forms especially, are not so convenient as the book-form slide. The shutters are made of hard rubber, ebonite, metal, or " compressed fibre," and draw right out of the slide for exposure, and in good makes care is taken to provide efficient light valves so that no harm occurs to the plates when this is done. The solid single and double slide is also made with flexible " roller blind " or " venetian blind " shutters (narrow strips of wood in close contact mounted on fabric). A prejudice has been shown towards these by some photographers, for which there is little or no justification, for they met the exacting conditions of aerial photography during the war when slides were used. For plates larger

[1] See " Light Filters."

[2] See Lees, S., " On the Superposing of two Cross-line Screens at Small Angles and the Patterns obtained thereby," *Mem. Manchester Lit. and Phil. Soc.*, 1918-19, lxiii.

than $8\frac{1}{2} \times 6\frac{1}{2}$ in., and for standing the wear and tear of everyday photography, the well-made double dark slides in sound mahogany would probably be preferred, because they are stronger and more reliable for "register." Good double slides for small plates are now made in metal, and small "single" metal slides are available, and both of these are particularly useful for hand camera work where it is desired to reduce bulk and weight in carrying cases. A refinement and a useful one in many slides is a safety catch or indicator to obviate double exposure on plates. For cameras used indoor, single slides of strong construction are usually employed, their greatest use being in copying and studio cameras. For copying cameras "roller blind" shutters are almost entirely used, and the same applies with modern studio apparatus save for small sizes. When a slide is to be employed for smaller plates than the maximum, inner frames or carriers are used, but "nested" carriers should be eschewed as they are liable to errors in register. In large slides, which are often built on a metal framework, carrier bars adjustable for different sizes are often employed. The essential requirements in a plate holder are absolute light tightness under the conditions in which it is used and accurate "register," which means that when the slide is in the camera the sensitive surface must be in the same plane as that previously occupied by the focussing screen when the image was focussed. The slide should be easily attached to the camera back, and the junction between the two must be sufficiently close to prevent the entry of extraneous light.

Cut sensitive coated celluloid films may be held in steel sheaths or carriers, which take the place of the plate, and adjustment for register must be made.

(ii.) *Changing Boxes.*—For holding a number of plates magazines termed "changing boxes" are often used with small hand cameras, but their use is not always so confined, for changers of different forms are largely used in aerial cameras, and some forms have proved satisfactory. One of the most reliable for use with the ordinary hand camera is the simple bag changer. In this the plates are held in a box in sheaths (preferably steel) and arranged one behind the other, the front plate being kept up to the "register" position. After exposure the plate is lifted up by a simple device and enters the bag at the top of the box, and is then grasped—through the bag—at the edges and placed at the back of the pile. The whole operation is quite simple, and where the boxes are well made the system is reliable.

In semi-automatic changing boxes the flexible bag is replaced by a drawer fitted with spring detents, and the action of transferring an exposed plate from front to back of the magazine is effected mechanically by simply pulling out the drawer (when the plate is projected into the lower part of the box by springs or falls by gravity) and then replacing it, when a fresh plate is ready for exposure. The device is fitted with a curtain shutter of wood or flexible shutter of metal, which protects the face of the front plate when the drawer is withdrawn, and a pull-out shutter or curtain which performs the same function as the shutter of an ordinary dark slide. An index is provided which shows a progressive number each time the drawer is pulled out, indicating the number of plates changed. It is necessary to note the number of exposures made in one form of box, as the changing action does not stop at the last plate of the series, but repeats; but in another form the last sheath of the series differs from those preceding and the drawer will not operate, thus indicating the completion of the series.

The ease and reliability of the changing boxes of this type, and indeed all changers, depend on good design and workmanship in the sheaths, and the boxes will not function when they are deformed, especially in mechanical changers. A well-known changing box as described is the "Ernemann" (Ica Co.), which operates by springs. This box may be used for small or medium-sized plates. Up to 5/4 the type works with twelve plates; above this size it is desirable to reduce the number, as the weight of the greater number is a bar to easy changing.

Changing boxes of the gravity type are well shown in the "Richard" (Jules Richard) and in the "Jacquet" (Tiranty). In this case the plates are small, the boxes are made with precision in metal, and the system is seen at its best. The experience of years shows that with reasonable care they are reliable.

All changing boxes require care in usage, and some operators are temperamentally unable to observe the necessary caution, hence the principal difficulties which arise in their use. As to whether or not they are desirable must be decided on considerations of general convenience.

§ (6) Panoram Cameras.—It frequently happens that photographs are required embracing a large angular "stretch" of country. Several negatives may be made by rotating the ordinary portable camera on the head of the stand, and prints from these negatives may be mounted edge to edge. Such pictures, however, will not join accurately. If, however, the camera be so mounted that the axis of rotation is vertically under the node of emission, this fault disappears, and cameras so mounted—which is quite simple—have been frequently used. Panoramic pictures may be made on film curved to the arc of a circle, the image being formed by means of a lens mounted at one end of a tube which has a narrow slit opening at the other end and coming close to the film. The tube swings through the arc of the circle about the

node of emission of the lens. The image is therefore stationary and in focus at all positions, and, provided the movement is sufficiently rapid, the camera can be held in the hand. This principle is utilised in the "Panoram Kodak" (Kodak), two models of which are available, giving angular pictures embracing 112 and 142 degrees respectively. A further development is the "Cirkut" camera (Century Camera Division, Kodak). In this type a mechanism unrolls flexible film past a slit opening, and by the same means the camera is caused to rotate on the head of the tripod stand. Negatives may in this way be made from 5 to 16 in. high, and for various lengths up to 16 ft., according to the camera and the length and width of the film used.

§ (7) Finders.—These are devices used for the purpose of indicating the amount of subject which would be included on the focussing screen of a given camera placed in the same position. In consequence, they may be employed in lieu of the focussing screen when the ordinary method of image adjustment by use of the screen is not practicable, and further they are required in the photography of moving objects. They are essential with all types of hand camera (save the Reflex, the focussing screen of which renders external agency unnecessary), with kinematograph and aerial cameras.

Finders are employed attached to the camera (or may form part of it), but finders may be used separately for preliminary observation, and are then, strictly speaking, to be regarded only as "view meters." The simplest device to use as a finder is the d.v. view meter, which is a small metal frame fitted with cross wires and a centralised rear sight, the sides of the frame having the same proportions as the camera plate, or is sometimes the actual size. The sight is placed at such a distance that view angle (plate) equals the field angle (subject), which distance may be determined by calculation or trial. The eye placed at the sight sees the subject through the metal frame, when correct pointing of the camera can be effected. In practice the eye is usually placed a little further away; this narrows the angle and gives a margin of safety, useful when moving objects are being photographed. In lieu of the wire frame there is sometimes employed a double or plano-convex lens with a mask showing the plate proportions, which requires a fixed sight correctly placed. In practice it is difficult to keep the eye fixed, and these finders are not reliable.

The instrument most frequently employed is the "brilliant finder," which consists generally of a little box with a small lens in the front, a mirror fixed at an angle to its axis, and a horizontally placed lens which acts as a magnifier to the image produced. The view is limited by a mask having the proportions of the plate, which mask is sometimes rotating. Other forms of "brilliant finders" have been introduced, but the one indicated, if well made and accurately adjusted, answers the purpose.

When the front of the camera is raised or lowered the ordinary finder as described is no longer reliable. To meet the new conditions the front supporting the image lens is made to raise and lower, and is in linkage with the rising and falling front of the camera, so that variation in the camera front automatically shows in the finder ("Identoscope"), whilst in other systems engraved scales are provided on finder and camera front, so that after arranging the view on the finder, which may mean raising or lowering its lens, the scale reading indicates the setting of the rising front to correspond ("Una") and ("N. & G.").

§ (8) Exposing Shutters. — These are mechanical contrivances for exposing the sensitive plate in the camera, generally, for brief durations. Several types are in use, of which the chief are (*a*) roller-blind, (*b*) focal plane, (*c*) between lens, and (*d*) flap. Both *a* and *b* are roller-blind shutters, but *a* works at the lens, whilst *b* is used in front of the plate. The roller-blind shutter is a flexible opaque blind, in which there is a rectangular opening travelling vertically across the front or the back of the lens, the movement being caused by a spring-actuated roller, the tension of the spring determining the speed of rotation, and in consequence the duration of the exposure. The efficiency is increased by increasing the length of the opening—which should never be less than twice the lens aperture—but this requires an increase in spring tension as compared with a shorter length to secure equal exposure duration, and high spring tensions are conducive to vibration. The blind opening must be wide enough to completely uncover the lens during exposure, and the framework of the shutter must not act as an obstruction, otherwise the margins of the plate will receive less illumination than the centre. When the lens is hooded the hood should be removed and fitted on the mount as close as possible to the front combination, otherwise an unduly large shutter is required. This shutter is shown in *Fig.* 10. The focal plane shutter consists of a moving blind in which is a rectangular adjustable opening—a "slit"—that is slightly greater in length than the longer dimension of the plate. The blind is actuated by spring roller, and when released travels across the plate. The duration of the exposure at any part of the sensitive surface depends upon the width of the slit opening and the rate at which the blind moves. The shutter is built into the

back of the camera, which must be enlarged for the purpose. The construction should permit the blind being as close as possible to

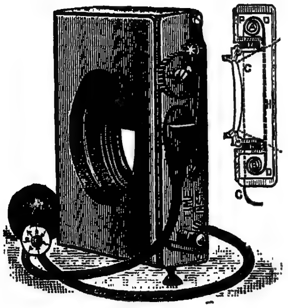


Fig. 10.

the plate, otherwise the efficiency is reduced. Two types are in use. In one the adjustment for exposure, so far as concerns one determinant, viz. the " slit," is effected by removing the blind from its channel and altering the width of opening by varying the length of the side cords or chains holding the parallel edges apart. The tension of the spring roller is then adjusted. With this form of the shutter (*Fig.* 11) care must be taken not to open the shutter of the dark slide before the blind is wound, otherwise if the lens be uncapped the plate will be accidentally exposed. To obviate this difficulty and to simplify setting, "self-capping" shutters are employed. In this form the slit in the blind is closed until the shutter is wound up so that the plate is protected from accidental exposure. The slit width is determined by an external setting device, the effect of which comes into operation when the blind is at the end of the wind. After exposure the slit automatically closes, and is ready for rewinding for a fresh exposure.

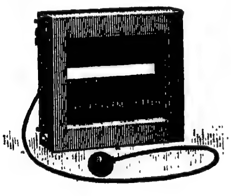


Fig. 11.

Focal plane shutters are necessary for very brief exposures. They suffer, however, from many defects — partly inherent, partly constructional. A serious defect is that the images are frequently distorted, due to the fact that the whole plate is not exposed at once but in sections, and although these defects are not always obtrusive they nevertheless exist.

The " between lens " shutter is generally a spring-worked sector diaphragm which opens from and closes to the centre, not interfering with the ordinary diaphragm for reducing aperture. The shutter mount replaces the ordinary mount of the lens. Such shutters are very convenient in use, but their efficiency is sometimes low, and their marked speeds are frequently not reliable.

Flap shutters are described by their name. They are used as a convenient substitute for the cap for exposures of appreciable duration. A flap—generally a light frame covered with velvet—is fitted in a mount which is fixed, generally, behind the lens. By means of a lever actuated pneumatically or by " antinous" release the flap is raised, uncovering the lens, and is lowered after the exposure by releasing the pressure. They are sometimes made with the flap in two parts to open in the centre, which is the better plan. Such shutters are generally only used in the studio. As well as " flap " shutters, a form of sector shutter actuated by pneumatic device is employed for this purpose and is to be preferred.

For the methods of testing shutters, and the considerations of their efficiency, see "Shutters, Testing of Photographic," Vol. IV.

§ (9) Stands.—The support for the camera is an essential part of the equipment, and its selection is generally an index of the user's judgment. Field or portable cameras are usually supported upon stands of the folding tripod type, although when employed as indoor instruments they are often mounted upon the rigid stand—the commonly termed " studio " stand—which will, as a rule, be found more convenient. The portable tripod is made of wood or metal, and ranges in form from the massive constructions with mechanically tilting and rotating heads used for cinematograph cameras to the simple rule joint folding tripod, which folds compactly and serves a less exacting purpose. Whilst wood is the more satisfactory material for the portable tripod for ordinary cameras, metal may be used, but really good stands would be more costly. Excellent tripods are made in steel tubing for cinematograph cameras. The most suitable wood is good well-seasoned ash, but other woods are employed—oak, cherry, maple. Rigidity is the essential point, and to this other requirements must reasonably conform. Tripod legs should be long enough to permit of their extension to form a wide base without reducing camera height below average eye level. For convenience the " legs " themselves are made folding and twofold with one adjustable element—usually termed " threefold "—are satisfactory if well made, the strength of the members of the stand being proportioned to the *size* and weight of the camera to be carried. The " head " should be for reliability approximate in width to the base of the camera, and the terminals of the legs

should be fixed to this as far apart as practicable. Heads are frequently provided with a tilting top, which is useful for architectural photography, and with universal heads to enable the camera to be levelled when the stand is in any position. Small ball-and-socket supports are, however, unreliable, except for tiny cameras. Whatever form is adopted, camera and stand should feel as if one, and the system should particularly resist torsional stress. For cameras used in the studio or indoors, permanent stands of the rigid tripod or two-pillar form, with tilting table and full range of rise and fall, are employed, the portable tripod being inconvenient and sometimes impracticable. For indoor work of a technical character a simple rigid tripod with rising pillar support, fixable by friction clamp, answers quite well, but for portrait work in the studio the requirements are greater.

§ (10) Enlargers.—Enlarging is a general term applied to the process of amplifying photographic images. Any camera system when the camera extension is sufficient to provide for the greater focal conjugate, will suffice in principle and is frequently used in practice. Actually, the enlarging equipment of many photographers is provided by using the ordinary camera in a darkened room, the instrument with the negative in place of the focussing screen (for a simple demonstration) being supported with its back before an aperture in an opaque screen fitted to the window frame. Outside the window is fixed a white diffusive reflector at an angle of 45°, which serves to illuminate the negative uniformly. Opposite to the lens is arranged a board covered with white paper upon a simple easel. By adjusting the respective distances of lens to negative and board to lens a sharp image enlarged to a particular amount will be formed, and eventually this image may be caused to fall upon a sheet of sensitive paper.

In *Fig.* 12 an arrangement on this principle is shown. A carrier for the negative is fitted

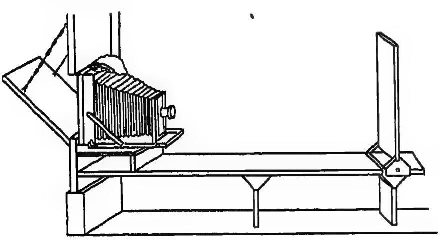


Fig. 12.

in the window frame, and the camera (with reversing frame removed) is arranged with its back close to the negative, extraneous light being excluded. The easel centred to the lens runs on a board as shown, and is so fitted as to be parallel with the negative at any distance. The apparatus is simple, inexpensive, and efficient.

Special apparatus is, however, designed for the production of enlargements, and its use involves fewer restrictions. Practical considerations generally require that the work be done by artificial light, at least in those countries where atmospheric conditions are inimical to uniformity, and, moreover, the necessity for providing for extended periods of work and at all times of the day and season require the use of sources of illumination additional to daylight.

The principle of the construction of the optical lantern underlies that of the artificial light enlarging apparatus, the elements being luminant, condenser, negative carrier, projecting lens, and screen, but in many instances an "extended source" of light is employed, in which cases the principle is simplified by the omission of the condenser.

A typical enlarging lantern comprises the "lamp-house" (containing the luminant), which is connected by bellows with the condenser body, beyond which is the negative carrier frame, and this, on the other side, is attached by a bellows to the front holding the projecting lens. The condenser is fixed, but the "lamp house" and lens front can be moved to and fro. The condenser and projector are mounted in line axially, and the luminant can be centred with these. Means are provided for raising and lowering the carrier, for its lateral movement, for swinging horizontally about a central axis, and for rotating it in a vertical plane.

The lantern as a whole may be conveniently mounted upon rails, upon which moves the screen easel, so that the two can be moved to and fro for adjustment of size of image, parallelism being maintained.

With respect to the luminant.

It is desirable that this, for optical reasons, be small in size, and to secure reasonably short "exposures" of the sensitive surface, intense. The small "enclosed" arc lamps, with right-angle carbon pencils, are convenient and efficient, but a "half watt" electric glow, with a suitable "concentrated" filament, or the tungsten arc ("Pointolite"), have many advantages and are more uniform in radiation. Acetylene, incandescent mantle (coal gas or spirit heated) are used with success. Vertical incandescent mantles often occasion difficulty owing to the large size of the illuminated area, and its frequent unevenness is a source of trouble. The inverted form is generally freer from these objections. It is often found that an image of the mantle appears on the screen. This may be avoided by placing between the condenser and source a piece of ground glass.

This then becomes the source, and rightly used should be provided with an adjustment to control the distance between itself and the light. From a purely optical point of view its use is to be deprecated. On the other hand, optical arrangements are frequently considered without taking account of purely photographic necessities. The image deposit in a negative is, as a rule, appreciably "grainy," and this causes "scattering" of the light, with the result that the gradations of tone in the enlargement are falsified as compared with those in the negative. Not only so, but flaws of a mechanical nature in the negative are often unduly apparent in the enlarged image, and the same applies to the marks of a spotting brush or retouching pencil. When the luminant is small these are, if not exaggerated, often unpleasantly visible; on the other hand, when the light is diffused, as when ground glass is employed, this objectionable appearance in the enlargement is not so manifest, so there is often a gain in departing from purely optical dictates. The ground glass is, however, something which upsets the action of the condenser when employed in the usual way, more particularly if placed on the side next to the negative (which is sometimes done), and this should be remembered.

When condenser systems are employed, it is necessary for even illumination of the negative that the diameter of the condenser be slightly greater than the diagonal of the negative, so that the whole area of the latter shall be included in the base of the cone of light. In practice this means placing the negative as close to the condenser as practicable. The focal length of the condenser should be suitable for the projector employed, or it may be found impossible even by adjusting the distance of the luminant to provide even screen illumination for every amplification desired.

If the negative is to be enlarged as a whole, and it be placed centrally, the raising, lowering, and lateral movements would not be required. But in practice it is frequently necessary to enlarge a portion only of a negative, and this not in the centre. It is desirable to place this area central to condenser and projector, and the movements named are wanted.

Tilting devices are for the purpose of correcting convergent distortion, which sometimes is produced in photographic negatives. For correct restitution similar arrangements must be provided to the screen receiving the enlarged image. Complete restitution of convergent distortion is a difficult and complicated matter, and the arrangements necessary are reasonably omitted from the ordinary enlarger, being required seldom and greatly increasing the cost.

The rising panel generally fitted to the front of the enlarger, by means of which the lens can be moved in a vertical plane, is not required, its use being wrong in principle, for the axis of the condenser and the axis of the lens should be one continuous line; moreover, it is quite unnecessary. The focussing adjustment for the front should be "fine," and it is a great advantage to have the means of focussing when the user is near to the screen. This is sometimes effected with enlargers having a central focussing screw (instead of rack and pinion at the side) by means of an extension rod fitted with Hooke's joint ("Ensign Premier," Houghton).

The projecting lens used is frequently that provided with the enlarger by the makers; equally it may be the one employed in making the negative, provided that the aperture be sufficient and the focal length suitable to the negative and the condenser—with daylight enlargers only the negative need be considered. The lens should be well corrected, and it is desirable that its principal focal length be not less than 1¼ times the diagonal of the negative to be enlarged. The aperture of the projecting lens permissible is largely governed by the size of the light source—with large sources, as an incandescent vertical mantle, a small angular aperture would mean inability to obtain an evenly illuminated screen. In practice it is desirable not to select a lens with an aperture less than F/6. There is no objection to the use of a lens having a greater focal length in relation to a particular negative than that specified, save on the ground of convenience. The increased distance required for large magnification may cause practical difficulties, and the ordinary condensers of commerce are not made for use with such objectives. In many respects increase of focal length would be advantageous.

With respect to screen or supporting sensitive paper, a plain board mounted upon an easel provided with the means for clamping on the rails so as to remain where placed is generally sufficient.

The board should be well framed to prevent warping, and may be covered with cork linoleum glued down, upon which a sheet of white paper has been pasted, or, better than the white paper, the linoleum can be painted with white "water" paint, renewed when dirty. There are, however, several special forms of easel which are convenient, and take various standard sizes of paper.

When a plain board easel is used a lens cap should be employed, glazed with a "safe" red glass or celluloid, and the paper is adjusted to the image as seen on the easel.

For making enlarged negatives a special easel, with carriers to take the plates, is very convenient, focussing being effected upon a

piece of ground glass used in the carrier. The enlarging lantern is frequently used for the purpose of making "reductions," *i.e.* pictures smaller than the original negative, as in lantern slide production. For this purpose it is well to have a special easel with a small focussing screen $3\frac{1}{4} \times 3\frac{1}{4}$ in., which swings out of plane when the image is focussed, when

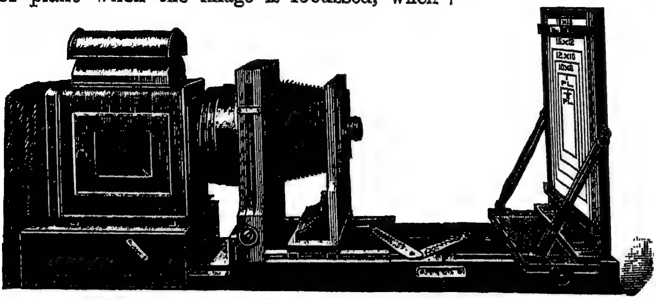


FIG. 13.

the sensitive plate contained in a similar holder takes its place.

Many efforts during recent years have been made to make camera systems, employed for copying and enlarging, "self-adjusting" or "self-focussing" or "automatic," either by setting to scales or, for the adjustment of the two focal conjugates, by means of cams and other devices. The apparatus is illustrated in *Fig.* 13 (Houghton Butcher). In this, focussing is effected by a device which in one movement adjusts lens to negative and screen to lens. To the lens panel is fixed a curved arm attached to one end of a right-angle cam, the other end of the cam terminating in a wheel that traverses a curved rail graduated to suit the focal length of the particular projector. The curved rail is fitted to an extension of the base of the easel, which is provided with a rack and pinion. As the easel is moved to and fro the wheel on the cam moves in the curved rail and adjusts the distance of the lens front. Focussing is, therefore, automatic, following the adjustment of the easel for the size of the image desired.

A further advance in this direction has been made by the introduction of the "Projection Printer" (Eastman), shown in *Fig.* 14. The apparatus consists of a camera supported above a screen upon which the image is projected. The camera unit consists of a "lamp-house" holding a 250 watt gas-filled metallic filament lamp, supported above the negative carrier, the sight area of the negative being regulated by means of adjustable masks. Below this there is arranged the projection lens, which is connected with the negative carrier by means of a bellows. The mechanism for "focussing" (using the term in its general sense) "consists of a cam and gear, the gear controlling the movement of the lens in relation to the negative, and the cam controlling the movement of the gear as the printer is raised or lowered to alter the size of the negative."

The swinging arms carry the camera unit, the balance being arranged that this portion of the apparatus moves up and down without difficulty, and as this is done the image varies in size, always remaining sharp. The paper is adjusted upon the screen below, and the margins are controlled by two right-angle masks. The light is controlled by foot-switch. To eliminate the defect in enlargements previously referred to, that of blemishes or marks in the negative being eliminated or reduced in the print, there is provided a series of optically worked glass diffusion discs which, placed over the lens, control the definition of the focussed image. This beautiful device is particularly valuable with portrait negatives which have been retouched, for with such

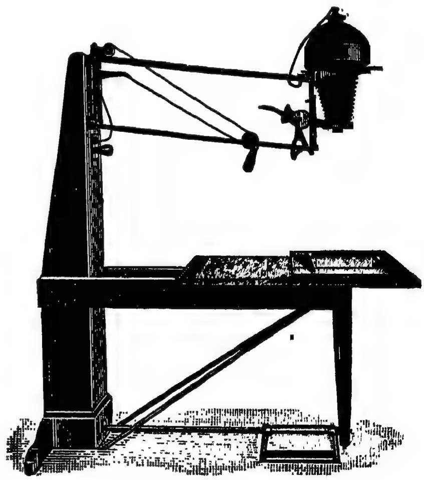


FIG. 14.

negatives there is always unpleasant prominence given in the enlarged print to the marks of the pencil in ordinary artificial light condenser enlarging systems. The apparatus takes negatives up to 7×5 in., and the range of amplitude from this size is from 10×8 to 56×40 in. There is a smaller size taking 5×4 negatives.

If the light source be much enlarged, so that it is greater than the negative, the condenser may be eliminated. This is the "extended source" system, and is now employed in many establishments very satisfactorily for the making of enlarged prints, particularly from big negatives. Enlarging apparatus constructed on this system use a brilliantly and uniformly illuminated area behind the negative, the light being either "transmitted" or "reflected." The initial source is generally a series of fairly closely placed straight mercury vapour tubes, or, for compactness, a specially constructed M-shaped tube. The light from this relatively large source is further diffused by sheets of suitably disposed ground glass. The system is shown in principle in the simple daylight enlarger shown in *Fig.* 12, the illuminated vertically arranged ground-glass screen taking the place of the opaque white diffusive reflector shown. In practice an indoor enlarging room is constructed with suitable movements for adjusting the lens to the negative and easel to lens. If an opaque screen be employed instead of the translucent arrangement it is generally illuminated from the front by carbon arc lamps or, for small equipments, by a series of gas-filled half-watt lamps, the negative being shielded from the direct light of the source employed.

There are many advantages, technical and economic, with an apparatus of this character. The light source requires no alteration for different degrees of enlargement—a matter of moment in industrial production—whilst with a condenser system adjustment must be made whenever there be a change of amplification. Enlargements made on this principle are far better as regards *photographic* quality, because the system of illumination by an extended diffused source is more suitable to the physical condition of the image in the negative, and, moreover, the enlargers are able to deal with large negatives which would be impracticable by means of condenser systems.

Extended source illumination enlargers—non-condenser systems—are becoming largely used in the form of vertical enlargers. Here the apparatus is fixed to the wall and occupies small space.[1]

Enlargements of small amplification, small box "fixed-focus" enlargers, are employed generally for daylight, and by these an enlarged print may be obtained by the simple operation of placing a negative at one end and a piece of sensitive paper—generally held in a dark slide—at the other, and exposing by turning the negative end towards the sky, or towards a white screen illuminated by daylight or on occasion by artificial light. More convenient enlarging boxes are made to work in conjunction with particular patterns of hand cameras for which they are designed.

§ (11) Hand Cameras.—The hand-held camera implies "exposures" of brief duration, and these have been rendered possible by the speed of the modern sensitive surfaces, and the production of satisfactory images by means of lenses of large angular aperture, with the subsequent elimination of the tripod and all that its use implies. Convenience dictates the employment of apparatus compact and easy to carry, simple in operation and quickly made ready for use. The employment of small plates from which enlargements could subsequently be made has reduced the necessity for large direct work, and the use of rollable film has caused photography to appeal to an ever-widening circle.

It is difficult, unfortunately, to find any form of classification for hand cameras which shall be really helpful and have, at the same time, the advantage of a logical justification, for hand cameras are the most individualistic things amongst photographic apparatus. The term "hand" camera is not in itself a distinction, and certainly not an exclusion, for any hand camera can be used as a stand camera if the operator so desires. The most distinctive introductions, so far as concerns type, have probably been the "Eclipse," which was the parent of one important class of folding camera in use to-day, the "Reflex," and in respect to far-reaching effect, the folding "Kodak," which embodied the idea of rollable film-holder and camera in one.

Modern hand cameras are, essentially, small picture cameras, the sizes ranging from $2\frac{5}{16} \times 1\frac{3}{4}$ to 5×4 in., and occasionally $6\frac{1}{2} \times 4\frac{3}{4}$ in. or "half-plate." They are used for plates or cut films, or both, but when the film is rollable the camera is generally made for this purpose.

A consideration or general application is the method of focussing. Disregarding the reflex (and the now seldom used "twin-lens" type) camera the actual image to which the plate will be exposed is not seen prior to exposure. The amount of subject included and the position of any portion therein, say the principal object, is observed by means of an external instrument—a "finder" (*q.v.*). The focussing may be done by judging the distance of the object and then setting the lens for that distance by means of a scale provided. The minimum distance of lens to plate will be when the lens is set for an object at "infinity," but in practice this position on the scale is of little value save as a datum line, for the best definition is seldom required for objects at a considerable distance. If the camera be of the

[1] For special apparatus see "The Printol" (Brown), *British Journal of Photography*, April 23, 1920, p. 257. For simple form of this convenient type see d'Arcy Power, H., *British Journal of Photography*, 1909, p. 475, and 1916, p. 328; and *British Journal of Photography Almanac*, 1910, p. 583.

"fixed focus" type the setting of the lens should be for an object at the hyperfocal distance of that lens when used at its maximum aperture. This plan secures sharp definition, for the greatest depth of field and objects exceeding a certain minimum distance are as well defined as possible *with that particular lens.* When the lens is set this way the depth of field may be taken, practically, from about half the hyperfocal distance onward to "infinity." The minimum distance becomes less if the aperture be reduced and therefore the depth of the field increases.

The "fixed-focus" cameras are advisedly only used for small plates when lenses of relatively short focal length may be employed. If, as is generally the case, the angular apertures are small the hyperfocal distance is near to the camera. When, therefore, the lens is set for an object at the hyperfocal distance, the "near point" being half this distance and the "far point" being "infinity," the depth of field is well extended. In consequence, by conforming to certain conditions such apparatus has considerable usefulness. By the employment of supplementary lenses readily attached to the camera lens, objects quite near may be brought into sharp focus. It should be remembered that hyperfocal distances are "extra focal."

In the majority of instances, when the camera has a focussing arrangement it involves the movement of the lens, either the front as a whole moves or the lens may be moved by a "focussing flange," the front remaining fixed. For accurate and easy working efficient scaling is necessary, more particularly as the negatives are frequently used for enlarging. It is desirable that a scale in *yards* be provided for the particular lens, so that when the index on the moving portion be set to any distance-mark on the scale the operator may know that an object at that distance will be "sharp" upon the plate. Focussing thus becomes a matter of judgment. But the production of the maximum sharpness for an object in one plane is only a particular case, although one frequently important in practice. More generally the operator desires to know the setting which will yield the best general definition for a series of objects situated within a certain range of distance from the camera, and for this "depth of field" scales are necessary. This form of scale can serve to some extent for both purposes, but since the numbers are not in simple sequence it is better to have in addition a separate scale in yards suitably graduated for what may be defined as "object focussing," for such work often entails rapid decisions, and any complication militates against quick conclusions. The two scales are now provided with some of the best types of camera, and can always be added to any focussing camera when desired.

In a reflex camera focussing is performed in the usual way. The image formed by the lens undergoes a directional change by means of a mirror, and is received upon a focussing screen at right angles to the plate. When the camera is held in the ordinary way the focussing screen is horizontal, and being suitably shrouded from extraneous light the operator looking down on the screen sees the image formed by the lens. The camera is therefore its own finder, and by actuating the focussing screen the image may be focussed as desired. The shutter release as a first action lifts the mirror out of place, and when the plate or lens is uncovered the image observed on the screen falls on the plate.

Practical considerations dictate compactness in bulk, and folding cameras are the most generally used. When the cameras are small the term "pocket" is used, but the size is the only criterion for this distinction.

A most useful form is when the camera opens concertina fashion, and the front and back are held apart by means of metal struts, or by side wings as in the parent form the "Eclipse." The front, when the camera is extended, is at such a distance that the plate is practically at the "infinity focus" or principal focal distance of the lens; any further extension required for focussing near objects is obtained by altering the distance of lens to plate by means of the "focussing flange" in which the lens is mounted. The exposing shutters are either focal plane or "between lens." This form is probably the best for rapid manipulation.

A type of hand camera follows the general style of the "light-form" portable field camera (*q.v.*), where the base board falls, and the front—to which the bellows are attached—draws out on the base, and focussing is effected by moving the front to and fro. Consider the focussing screen (where present) protected, and that the whole of the movements are enclosed when the camera is folded, and there results one of the most important forms of hand camera. Such instruments are provided with focussing adjustments, rising, falling and swing fronts, reversing back, focussing scales and finder. The exposure device is either a focal plane or "between-lens" shutter. These cameras are employed for plates, held in dark slides, or for cut films. Examples are shown in the "Una" (Sinclair), "Vesta" (Adams), "Sibyl" (Newman Guardia), and "Sanderson" (Houghton Butcher).

The typical camera for roll film is the folding kodak (Kodak, Ltd.). This apparatus, when closed, is self-contained, the sensitive material being held in a spool in the camera. The "base" on release falls and locks in a

horizontal position, and the front draws out, sliding upon parallel guides in the base, focussing being effected by means of rack and pinion. The essential movements are provided as in any other camera, but additions and refinements are provided in the many forms available. A distinctive feature of the kodak type is the employment of the "daylight-loading" spool of film, by means of which the camera can be replenished even under open-air conditions, a convenience which has had a most important bearing upon the diffusion of the practice of photography. The sensitive film is changed after each exposure by a winding key, and an indicator is provided showing when the change has been completed. The exposure is made by means of a "between-lens" shutter. In the "Kodak" camera known as "Autographic" there is provided the means to inscribe the title of each exposure upon the film before changing, and the developed pictures bear their own record.

The distinctive principle of the "Reflex" camera has already been described. This camera is advisedly the best for the more studied work of those advanced photographers with whom careful pictorial composition is the rule. Although good composition is determined by a visual study of the subject itself, it is undoubtedly an advantage to be able to see the image upon a screen, the same size as the plate, before exposure. Inasmuch as the essential requirements of a good camera remain constant, there is only the special addition to be considered. The ordinary reflex employs a focal plane shutter, it being necessary to protect the plate "held in waiting" from exposure until the mirror is lifted out of the way, the perfect closing of the top of the camera so as to prevent extraneous light from entering through the finder screen being essential. The drawback in many respects to the focal plane shutter is that it is not entirely satisfactory for the relatively slow exposures required in much reflex camera work, especially that of the pictorial kind. This drawback has been overcome in the "N.S. Reflex," where the focal plane blind has been eliminated and a "between-lens" shutter substituted. A supplementary flap protecting the plate follows the mirror when the release is pressed; this flap, however, only comes into action when the lens shutter, open for the focussing, has returned to the closed condition. By this device the plate—uncovered ready for exposure—is protected until the lens is covered, and the mirror having risen excludes any light which otherwise would enter the camera through the ground-glass finder screen. At this stage the exposure takes place. A most important necessity in the reflex camera is the perfect functioning of the release system which controls the sequence between the raising of the mirror and the exposure.

As an example of the reflex camera the "Soho" (Marion) may be cited.

The reflex camera is of necessity somewhat bulky and relatively heavy, but with the tendency to employ small sizes this drawback to the use of a beautiful instrument is mitigated. To overcome the bulk difficulty folding reflex cameras have been made, but it may be observed that the folding principle does not receive its best exemplification in the reflex camera.

As well as monocular cameras the stereoscopic principle is applied to hand cameras, and in these to-day it finds its principal use.

§ (12) Aerial Cameras.—The aerial camera, leaving aside the forms employed in survey during the last 60-70 years, owes its development to recent military necessities. As originally designed for the purpose of reconnaissance, the aerial camera was, essentially, a simple fixed focus box camera of wood or metal provided with a lens of large angular aperture, a focal plane shutter, and a simple finder, the plate being placed at the principal focal distance of the lens. The plate receptacle used was either a changing box or some form of single or double plate holder. The camera was hand held. The principal focal distance of the lenses employed was 8·25-10 in.

To avoid the many difficulties inherent in hand-held cameras, the apparatus was afterwards fixed externally to the side of the fuselage of the machine, or internally. Eventually the camera was fitted with a simple changing device. The plates held in steel sheaths were contained in a magazine, and were received after exposure in a receptacle of the same kind. The apparatus was operated by hand, the plate being changed and the exposing shutter set by one operation, exposure following at will by actuating the shutter release, and the cycle was repeated for the next plate. Subsequent developments led to a much improved changer system and to the building of the camera wholly of metal. The camera was also separable into two portions, the upper part or body forming the changer system with the shutter, and the lower—the cone—holding the lens. By the use of cones of different lengths it was possible to utilise lenses ranging from 4½ in. to 20 in. focal distance. The shutter was made detachable and of the self-capping type, with external adjustments for slit width. In this apparatus changing the plate and setting the shutter is performed by mechanism operated by hand or by motive power derived from a small independent propeller on the exterior of the aeroplane, driven by the current of air as the machine travels through space. The camera

can be used intermittently for single photographs or continuously for a series, determinable at will, so long as plates remain in the magazine. The camera mechanism can be put in or out of action when at a distance from the observer by means of a Bowden wire. To avoid the effects of vibration the apparatus is fitted on a special cushioning device. The negatives made in small cameras (5/4 in., 9×12 cm. and 13×18 cm.) may be enlarged, but, in lieu, negatives of greater size may be made in cameras of similar form to those indicated, using dark slides or magazine changers hand operated, or in automatically worked apparatus, all of which methods having yielded good results even under the relatively trying conditions of aerial flight. To avoid the use of plates, roll film cameras of different types, automatically driven, have been used. Attention has been paid in every country interested in aerial photography to the design of these cameras, and it is probable that in these the main line of development will be made in the future. One of the principal difficulties with film cameras is the want of planarity of the film. In one form of apparatus it has been sought to overcome this trouble by pressing the film against a plate of glass, and in other instances suction devices have been employed to maintain the film flat in the focal plane, and this plan offers probably the fewest drawbacks.

C. W. G.

REFERENCES

Aerial Cameras.

"Aerial Cameras," Laws, V. F., *Roy. Photo. Soc. Trans.*, 1920, lxix.

"Aircraft Photography in War and Peace," Hamshaw Thomas, H. (Cantor Lectures), *Roy. Soc. Arts J.*, 1920, lxviii.

"Improvements in Lenses for Aerial Photography," Appleton, W. B., *Roy. Photo. Soc. Trans.* lix.

"On Some Photographic Apparatus used in Aerial Photography," Gamble, Charles W., *Optical Society Trans.* xx. 163.

Les Applications de la photographie aérienne, Clerc, L. P., Paris, 1920.

Airplane Photography, Ives, H. E., Philadelphia, 1919.

For Kinematograph Cameras see "Kinematograph," Vol. IV.

PHOTOGRAPHIC LENS, MEASUREMENT OF FOCAL LENGTH OF. See "Objectives, Testing of Compound," § (2) (i.).

PHOTOGRAPHIC LENSES

§ (1) INTRODUCTORY.—The lens in photography is employed to concentrate a large amount of light of certain wave-lengths in the form of an image of external objects on a thin layer of sensitive matter which is otherwise shielded from all light. Convenience requires the sensitive layer to be in the form of a plane, and the exceptions to this rule may for the purposes of the present article be neglected. According to elementary theory a lens produces, for light of any given wave-length, collinear images of all points, and consequently the image of a plane is itself a plane. According to this theory any lens, if achromatic, will tend to give a plane image of say a picture, and this image will be an exact copy of the original picture if the axis of the lens meets the plane of the picture at right angles. By choosing a position for the lens at a suitable distance from the picture the image may be made as large or as small as is desired, and thus any lens which transmits a cone of rays wide enough to include all parts of the picture will be suitable for the reproduction of the picture to any size required. This conclusion has to be modified when aberrations are taken into account, but, except with lenses of very large aperture, it is sufficiently correct for most photographic work provided the lens is of good design.

The case of the plane object, which is almost invariably the only one considered in textbooks, is not the most frequent one in actual photographic work. The typical problem is the reproduction of a scene out of doors where objects at very varied distances from the lens are included within the angle of view. Some of these it is desired to have recorded in sharp focus on the sensitive layer though they may be at different distances from the lens, and again it may be desired that other objects shall, as far as possible, not be evident in the reproduction on the sensitive surface. Moreover, some of the objects may be in motion, and the time during which the surface may be exposed to light has therefore to be made small. Hence the aperture of the lens is required to be large in order to transmit sufficient light in small intervals of time. In many cases the image is required to cover a large solid angle also. All these considerations introduce special difficulties, some tending in their influence on the design of the lens in one direction, others in an exactly opposite direction. It will readily be appreciated that under the circumstances considerable diversity is called for in the design of lenses, and that even so each one represents a compromise between conflicting requirements, the aim of the maker being to balance the desirable qualities in such proportions that each lens he manufactures will be found sufficiently satisfactory by a large class of users.

§ (2) CENTRAL ILLUMINATION. (i.) *The Stop.*—However small they may appear to be, all the objects suitable for terrestrial photography must be regarded as of finite size, and it follows that in considering the illumination at any part of the image the finiteness of the size of the image must be taken into account. Suppose that the centre of the field of view is occupied by a small plane object of

area S placed normal to the lens axis. If the distance between the object and the lens is d and the area of the stop is A, the total amount of light from the source which passes through the lens will be represented with sufficient accuracy for our purposes by

$$\frac{\mathrm{ISA}}{d^2},$$

where I is proportional to the light given out by unit area of the object. If the lens is properly designed, all this light, with the exception of a proportion which is unavoidably reflected or scattered when the light is refracted at the surfaces of the lens, will reach the sensitised surface. If the area of this image is S′, the illumination on unit area of this surface will be

$$\frac{\mathrm{ISA}}{\mathrm{S}'d^2}. \quad . \quad . \quad . \quad . \quad (1)$$

If the stop is so placed that its images are in the unit planes, that is to say, if the stop image appears of the same size when examined from the front of the lens as when seen from the back, and d' be the distance between the second unit plane and the image of the object, then, since the magnification m is measured by the ratio of the distance of the image to the distance of the object, we have

$$\frac{d'^2}{d^2}=m^2=\frac{\mathrm{S}'}{\mathrm{S}}.$$

It follows that the brightness of the image is proportional to

$$\frac{\mathrm{IA}}{d'^2},$$

or

$$\frac{\pi\mathrm{I}}{4}\left(\frac{a}{d'}\right)^2, \quad . \quad . \quad . \quad (2)$$

where a is the diameter of the stop when this is circular. When the object is fairly distant d' differs by a negligible quantity from f the focal length of the lens, and the exposure required to produce a satisfactory impression on the sensitive surface is therefore proportional to $(f/a)^2$. The sizes of stops are most generally denoted by numbers which indicate the values of f/a, and it is customary to designate the rapidity of lenses by quoting the smallest value of this fraction for which the lens can be used. Thus a lens may be described as having the rapidity $f/4$, meaning that the diameter of the largest stop it is possible to use is equal to one-quarter of the focal length. Such a lens would only require an exposure of one-quarter the time necessary with one in which the largest stop is $f/8$, or one-eighth of that wanted when a stop as small as $f/11$ is required.

(ii.) *The Effect of Distance.*—When the object considered is not very far away the image is removed to a greater distance from the lens, so that d' increases; the image falls off in brightness though the object remains equally bright. This is partly the reason why in a photograph of a scene in which both near and distant objects are included the former appear unnaturally dark. In copying a plane picture on any given scale the increase in the time of exposure required is independent of the focal length, since under these conditions the ratio S : S′ is given and d and d' are proportional to the focal length. In the case of natural objects the result, however, is dependent on the focal length, and our standard of comparison is provided by the effects we see with our eyes, instruments having a very short focal length. By a very well known optical formula the linear magnification for an object distant d from a lens of focal length f is

$$\frac{f}{(f-d)},$$

and therefore (1) becomes

$$\frac{\pi\mathrm{I}}{4}\left(\frac{a}{f}\right)^2\left(1-\frac{f}{d}\right)^2, \quad . \quad . \quad . \quad (3)$$

and the illumination falls off the more rapidly the greater the value of f. Taking 15 mm. (0·6 in.) as the focal length of the eye, the variations in light intensity due to proximity will be at least 5 times as great with a lens of 75 mm. (3 in.) focal length as with the eye, and at least 20 times as great with a lens of 20 cm. (12 in.) focal length. This characteristic indicates that under certain circumstances there may be a considerable gain in employing a lens of short focal length and enlarging the resulting picture rather than in taking a view directly of the size ultimately required.

(iii.) *Scattered Light.*—It has been mentioned already that a certain amount of light is inevitably scattered at each refraction. The amount so lost is far more detrimental to the production of brilliant images than the subtraction of the corresponding amount of light from the incident beam would indicate, for some of the scattered light from intensely illuminated objects will fall upon parts of the image corresponding to dark objects, with the result that a flattening in the tones of the image takes place. It is important to avoid this as much as possible, and it follows that lenses which have a small number of glass to air surfaces tend to produce brighter pictures, as the amount of light scattered is almost entirely due to refraction to or from air, that dispersed at surfaces where two glasses are cemented together being in comparison negligible.

It should not, however, be supposed that the presence of a number of such glass to air surfaces is the chief cause of flat photographs. Far more serious sources of trouble in many cases are due to carelessness in allowing bright

light from a source outside the proper field of view to fall upon the interior of the lens mount on the one hand, and to an attempt to secure undue compactness in the apparatus on the other, which leads to the omission of a lens hood, and still more serious, to the use of too small a bellows to the camera, causing a great deal of light to be reflected from the bellows on to the plate. A proper bellows should be large enough to allow two or three flexible stops to be inserted which will intercept light that might otherwise cause general fogging.

§ (3) Illumination in the Outer Parts of the Field.—Suppose now there is in the outer parts of the field of view another area similar to that at the centre, of the same brightness and in the same plane. Let the angle made with the axis by the line joining the centre of this area to the centre of the stop be ϕ. Suppose the object behaves as a self-luminous body. Then the amount of light emitted per unit area in a direction making an angle ϕ with the normal to the surface is proportional to $\cos\phi$, so that I must be replaced by $I\cos\phi$. The stop seen from the object appears foreshortened, its area being apparently $A\cos\phi$ instead of A. This substitution must therefore also be made. Further, corresponding to the distance d we have the increased distance $d\sec\phi$. S and S′ are of course unchanged. Taking all these factors into consideration, it appears that the illumination in the image of an evenly illuminated object falls off in proportion to $\cos^4\phi$. The accompanying diagram (*Fig.* 1) exhibits this graphic-

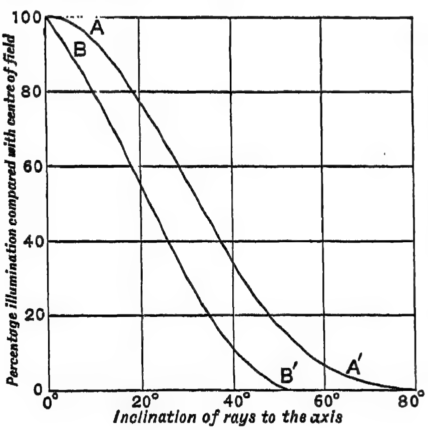


Fig. 1.

ally in the curve AA′. It will be seen that when

$\phi=20°$	the	illumination	is reduced	by	22	per cent.
$\phi=25°$	,,	,,	,,		33	,,
$\phi=30°$	,,	,,	,,		44	,,
$\phi=35°$	,,	,,	,,		55	,,

Serious as these figures are, they by no means represent the worst, for it almost invariably happens that for object points only a small distance from the axis a second limiting aperture begins to cut down the light still further below the figures given in this table. The way in which this comes about is illustrated in *Fig.* 2. A central beam is shown

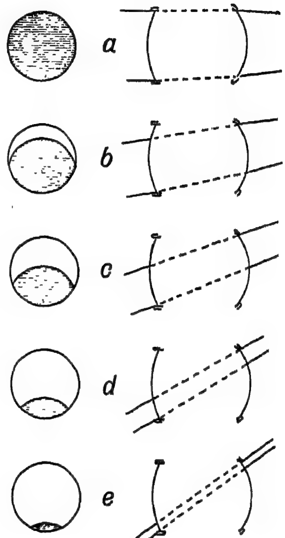


Fig. 2.

passing through the lens in section at *a*, a beam at 10° at *b*, at 20° at *c*, 30° at *d*, and at 35° at *e*. The reduction in cross-section is evidently due to the axial separation between the extreme lenses, and its influence on the illumination is indicated in the curve BB′ of *Fig.* 1. It is quite common to find that the illumination at 30° is of the order of 15 per cent of that at the centre. In spite of the 7 to 1 ratio in light intensity which this represents it is not usually noticed in photographs, because the eye is not very sensitive to *gradual* changes of intensity. If, however, the angle is made unusually large, such as by a very liberal use of the rising front of the camera, or if a composite photograph is made with "straight" prints, the variations are noticed immediately.

When it is necessary to cover a very large angle the most ready method of securing the greatest uniformity in exposure is to prevent the secondary reduction of aperture by stopping the lens down until the whole circle of the aperture can be seen from the corners of the image it is required to record. In some lenses the stop is so placed that this reduction is of no effect, and in a few cases stopping

down actually reduces the size of the total field of view and increases the disparity in illumination. Quite frequently it is necessary to reduce the size of the stop for this purpose to a far greater extent than is consistent with very sharp definition in the image. When this expedient is insufficient, recourse must be had to special stops contrived to reduce the amount of light at the centre to a greater extent than that at the edges. The principle employed consists essentially in the combination of a circular opening in a plane with an opaque circular stop in another plane, or alternatively with a conical stop (*Fig.* 3).

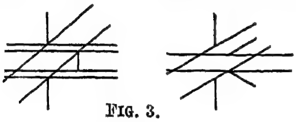


FIG. 3.

These arrangements date from a very early stage in photographic history, and were developed by Bow and Sutton. At the present time the best known application of this principle is in the case of the Goerz "Hypergon," in which a rotating star-shaped central stop is used. With this lens ϕ may be as great as 70°.

§ (4) DEPTH OF FOCUS.—When the view contains objects at different distances from the lens, it is not possible for the images of all of them to be formed in the sensitive layer, and the impression produced is that of a more or less badly focussed image. The simplest way of ascertaining the effect which will be produced is to neglect the depth of the sensitive layer, so that only one object plane is supposed to be sharply focussed. If P is an object point outside this plane, as at P_1 or P_2 (*Fig.* 4),

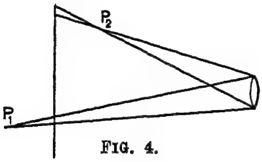


FIG. 4.

a geometrical theory leads to the conclusion that P, instead of being represented in the image by a point, will appear as a patch corresponding to the image of the circle in which the cone, whose vertex is at P and whose base is the lens aperture, intersects this plane. If d is the distance of the plane from the lens, and d_1 is the distance of the plane through P parallel to the plane rendered sharply, the diameter of the equivalent object circle will be ϵ, where

$$\frac{\epsilon}{a}=\frac{d_1 \sim d}{d_1}.$$

It follows that if the nearest and farthest objects to be rendered in acceptably sharp focus are distant d_1 and d_2, the proper distance to focus on is d, where

$$\frac{2}{d}=\frac{1}{d_1}+\frac{1}{d_2},$$

and the diameter of the circle in the image plane which serves as an image of a point in either of the two extreme planes is

$$\frac{a(1/d_1 \sim 1/d_2)}{2/f-1/d_1-1/d_2},$$

where f is the focal length of the lens and a the diameter of the stop in use.

In applying this formula it is usually assumed that a blurred image of a fixed diameter may be accepted as a standard, and the most usual tables adopt ·01 in. as this diameter. Probably the assumption of a fixed diameter is not unreasonable for lenses of focal lengths 8 in. or more, but this criterion appears unsuitable for photographs taken with lenses of very short focal lengths such as are now common, where it is reasonable to assume that every picture must be enlarged to be properly appreciated. In these cases the diameter of the blur should probably be reduced in proportion to the focal length, attaining the value for larger lenses when the focal length reaches approximately 8 in. For securing a sharp impression there is no doubt that in the case of the longer focal lengths ·01 in. is too large a value to assume, at any rate for objects of importance in the central parts of the view, and to attain the standard reached by the normal eye this requires to be reduced to about one-quarter. On the other hand, to make sure that certain objects will not be distinguishable in the print, it is desirable that a value exceeding ·01 in. should be adopted.

One of the most important applications of the above theory concerns the focus for "infinity" in small hand cameras. Here, if anywhere, the ·01 in. limit is tolerable, as the principal objects in views taken with such cameras are rarely really distant, and it is only important to ensure that the blurring of very distant objects does not present an unnatural appearance to the eye. For this case the formula becomes

$$\frac{\text{Distance to object sharply focussed for infinity}}{\text{Focal length}}$$
$$=1+\frac{\text{Diameter of stop}}{\text{Diameter of blur circle}}.$$

Thus with a lens of 5 in. focal length and full diameter 1 in. the object sharply focussed when the camera is set for infinity will be slightly over forty feet away, and the sensitive surface will be about ·05 in. from the plane conjugate to infinity.

§ (5) Depth of Focus and Phase Errors.—It is interesting to consider in connection with these traditional rules the results which would follow from the application of criteria for good definition suitable for the telescope, finding the depth of focus for instance in terms of the phase errors consequent on departure from the theoretical focus. Considering a point on the axis from which a ray to the rim of the lens aperture makes an angle ψ with the axis, the difference in path introduced by changing the distance of the point from the lens from d to d_1 for this extreme ray is

$$(1-\cos\psi)\,(d_1 \sim d),$$

and since $\sin\psi = a/2d$, this path difference may be written as

$$\frac{a^2}{8}\left(\frac{d_1 \sim d}{d^2}\right),$$

or

$$\frac{a^2}{8}\left(\frac{1}{d} \sim \frac{1}{d_1}\right).$$

If, then, $\pm c$ represents the greatest path difference consistent with acceptable definition, the connections between a, c, the near and far distances d_1 and d_2, and the distance d on which the lens should focus are

$$\frac{2}{d} = \frac{1}{d_1} + \frac{1}{d_2}$$

as before, and

$$\frac{1}{d_1} \sim \frac{1}{d_2} = \frac{16c}{a^2},$$

a formula differing markedly from that obtained by considering the area in which the ray bundles meet the image surface. According to this theory the depth of focus is inversely proportional to the area of the aperture, instead of varying inversely as the aperture diameter and the camera extension. The connection between the quantities involved in the two theories may be written

Path difference

$$= \tfrac{1}{4}\text{ diameter of image circle} \times \sin\psi_0,$$

where ψ_0 is the angle made with the axis by the extreme ray in the image space which passes through the axial point of the image plane. Taking an object fairly distant from the lens on which to focus, and assuming an aperture of $f/4{\cdot}5$ and a phase tolerance of 1 wave-length of yellow light, the diameter of the permissible image circle is found to be approximately ·0007 in. $=1/1400$ in. On account of the finite thickness of the sensitive layer and the appreciable size of the grains which become developable as a whole when affected by light, it is doubtful whether so small a tolerance is necessary under ordinary conditions. The wave-length theory is, however, undoubtedly applicable in considering to what extent accuracy in focussing has any meaning. Taking, as before, the case of a lens focussed on a distant object, it would be expected that under the best conditions imperfection in focussing would begin to be visible when the image plane was out of focus by about ·001 in., and for other stops the distance would increase as the square of the f number, and so be proportional to the correct exposure. Thus, with an aperture of $f/8$ the focussing depth in the present sense would be three times that for $f/4{\cdot}5$.

§ (6) Tilted and Decentered Use of Lens.—It is evident from the foregoing discussion that when it is desired to record in good focus two objects in different parts of a view which are at very unequal distances from the lens, the latter must either be used with a small aperture, or alternatively the lens must be used with its axis inclined to the normal to the surface on which the record is to be made, so that both objects lie approximately in the plane conjugate to this surface. Under these circumstances the plane containing the sensitive layer and that in which the objects are situated will be inclined to one another, and it is of importance to be able to make the necessary adjustment with facility. This is almost invariably carried out by swinging the back of the camera, with which the sensitive surface moves, relatively to the lens axis, which is pointed as usual to the centre of the view it is intended to include. The proper adjustment is indicated by a collinear theory of imagery. Since the object plane is not normal to the lens axis it will meet the first unit plane of the lens in a certain line, and hence the image plane must meet the second unit plane in a corresponding line. With almost all lenses the unit planes are very slightly separated from one another, and lie between the extreme components of the system. It is sufficiently accurate for the purpose of adjusting the camera back to assume that there is a common unit surface which bisects the lens mount. By considering where this plane is met by the plane it is intended to have in sharp focus, and directing the image plane through this line the correct approximate adjustment is easily made, and if necessary a final correction can be made after the image has been examined. The principle followed is illustrated in the accompanying figure (*Fig.* 5), where VQAP is the object plane meeting the common unit surface VC in V, and Vpaq is the image plane.

Obviously from this figure the view will not appear correct in the final picture unless it is seen from one side, or, if the object and image planes were inclined to the vertical, from above or below. In certain cases this effect is very objectionable, and in particular it must be avoided in architectural subjects where the use of a lens with its axis inclined

to the horizontal causes vertical lines in the object to appear to converge in the image. The simplest method of dealing with such

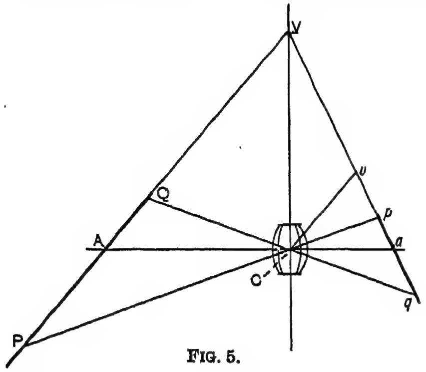


FIG. 5.

subjects when, as frequently happens, the desired view cannot be obtained with the axis of the lens intersecting the centre of the sensitive surface is to decenter the lens with respect to this surface, so that, for instance, in taking a view of a tall building from the ground the lens axis may be several inches above its normal position. The parallelism of vertical lines is thus preserved in the picture and the whole face of the building is kept in satisfactory focus if the image plane is kept vertical. This method, in extreme cases, is, however, unsatisfactory on account of the poor quality of definition in the image in parts remote from the lens axis, and also on account of the extreme difference thus introduced in the effective exposures for different parts of the image. In these extreme cases it becomes necessary to use the lens in an inclined position, and to rectify the convergence of vertical lines by re-photographing the original record.

Referring again to the figure, which shows a section in the plane containing the axis and the projection of the axis on the object plane, let Cv be a line, parallel to the object plane PQV, through C, the axial point of the common unit planes CV, meeting the image plane Vpq in v. Then v is the vanishing point in the record for the converging lines corresponding to the parallel vertical lines of the object. In copying this record so to correct this convergence it is evidently only necessary to ensure the parallelism of the line of greatest slope of the new image plane to Cv when pq is treated as the object. In order, however, to obtain a correct relation between horizontal and vertical distances in the final picture, the axis of the lens in this second operation must intersect the first record now serving as an object in the same point a as did the axis of the lens when the record was made. If the final record is to be increased in scale to m times that of the first record, the unit plane in the principal section must be made to meet the record in V where $\mathrm{V}v = m \cdot va$. If now a circle is described on Va as diameter the axial point of the unit plane C must lie on the circumference, thus making VCa a right angle. It is evident from this construction that the only limitation which must be imposed on the lens used for the second operation is that its focal length shall not exceed $\frac{1}{2}\mathrm{V}a$ or $\frac{1}{2}(1+m)va$ where m is considered positive. The original lens is, in such a case as has been considered, suitable for the purpose, but one of shorter focal length is frequently more convenient in order to reduce the axial separation between the object and image planes.

§ (7) THE POSSIBLE DEFECTS OF PHOTOGRAPHIC LENSES.—The matters that have been discussed hitherto relate to the correct use of a lens, and apply to all lenses alike, whether they are well or less well corrected for aberrations. The defects now to be considered depend upon the design of the lens, and are not under the control of the user.

(i.) *Filming.*—A defect which is not likely to be present in a new lens, but may develop in the course of time, is the filming of one or more of the glass surfaces. This is particularly likely to occur with extreme barium crown lenses of the types used in anastigmat lenses. When this filming is slight it is innocuous, but in some cases it causes the lens to appear of a decidedly brown tinge when examined by transmitted light, while iridescent colours may be seen by reflected light. In these more extreme examples the rapidity of the lens is seriously affected, and it is necessary to have the lens surfaces re-polished. It is needless to say that this should only be done by the maker of the lens.

Somewhat similar colours may be seen by reflected light when two surfaces which are supposed to be cemented together have become partially or entirely uncemented. The most probable cause of this defect is a fall or violent shaking. In these cases also the lens ought to be returned to the maker for correction.

(ii.) *Mounting Defects.*—A lens that has been performing satisfactorily may be found unsatisfactory after cleaning or re-mounting. This is frequently experienced after a lens formerly carried in the maker's mount has been fitted into a between-lens shutter. The cause may be a scratch on the lens surface, which scatters light in all directions within the camera; strain due to screwing up a lens too tightly in its cell; changes in the separations of components either through a lens being loose in its cell or to incorrect adjustment of the new mount; the insertion of a particular component the wrong way round, or to interchange of front and back com-

ponents; or to inexact adjustment of the axes of the various components into coincidence. The accuracy necessary in some of these adjustments is very often not realised, and particular care to see that all is in order is essential after any disturbance of the lenses from their original position. Even when all the parts are in relatively correct positions the optical axis may be found appreciably inclined to the mechanical axis of the screw by which the mount is attached to the camera front. The figures already given on the accuracy desirable in focussing indicate the great nicety to which these two axes ought to be brought into parallelism if satisfactory records are to be obtained.

(iii.) *Faults of the Glass.*—Defects of a different class are faults in the glass of which the lenses are made. Bubbles are practically unavoidable in some of the new types of glass which are always used in anastigmatic constructions, but they should not be too numerous or too large. On these conditions they are practically harmless, their chief effect being the stoppage of a small amount of light. More serious are veins and striae. These can only be seen under special conditions of illumination if they are slight. The simplest arrangement is to point the lens at a small source of light, and place the pupil of the eye at the image of the source formed by the lens. The defects will then become apparent as lines where the illumination is not uniform. Lenses showing such defects cannot be made satisfactory, and should not be accepted as good instruments. In the physical laboratory lenses or prisms showing these defects can usually be discovered, and are interesting subjects for experiments with the aid of an interferometer and in other directions.

(iv.) *Defects of Design.* — The remaining defects of importance depend upon the design of a lens. It has already been mentioned that light is reflected and scattered whenever a refraction from glass to air or *vice versa* takes place. For this reason, other things being equal, the design with the smallest number of such refractions is to be preferred. Another defect which may be coupled with this is the liability of a lens to show flare images. These images are images of the lens aperture formed by light that has been reflected twice at glass to air surfaces. As a rule these images will not be situated near the image of the object to be recorded, but in certain designs one or more of these images may occupy a position where it is clearly seen in the record. A lens having this defect is quite unsuited for photography under certain conditions of lighting. It is well known that one of the best known makers of photographic lenses was compelled to withdraw an otherwise eminently satisfactory lens from the market shortly after its introduction, owing to its marked failure in this respect.

§ (8) Aberrations.[1]—The remaining regular defects to which lenses are subject are of the character known to the optician as aberrations, and it is to the removal of these defects by the use of suitable materials and by the selection of suitable forms for the component lenses that the designer principally directs his attention.

(i.) *Chromatic Aberrations.*—The simplest of these defects are those known as chromatic aberrations; the images of the object formed by light of different wave-lengths differ in size and also occupy different planes. For any object position there are thus two primary defects to remedy, and if images at different distances from the lens are to be corrected three conditions must be satisfied. The crudest lenses fitted to some of the very small cameras are not corrected for these defects, advantage being taken of the very small size of the lens aperture which renders the fault less obtrusive in the picture. In all cases where a non-achromatic lens is used a visual examination of the image shows it to be out of focus, an improved appearance resulting from an increased separation between the lens and the image plane. This difference is due to the fact that the sensitive materials used to secure a photographic record are most sensible to light of a shorter wave-length than that to which the eye attaches chief weight, and the actinic or photographically active image is thus nearer to the lens than the visible image. The difference is of the order of 2 per cent of the focal length of the lens.

In more pretentious lenses this difference in the position of the image planes is always removed by using combinations of positive and negative lenses in which the negative lenses are constructed from glass having a greater dispersive power than that used for the positive lenses. The three chromatic conditions reduce to one when two thin lenses, one positive and one negative, are used in contact with one another. When the system consists of well-separated lenses, or of lenses of considerable thickness, attention has to be paid to the three conditions independently of one another, but as a rule it will not be found that the theoretical conditions are exactly satisfied, the designer finding it more convenient to reduce the defects to a point at which they cease to become evident than to eliminate them completely. In one important case the three conditions reduce to two through two of them assuming an identical form. This is when the complete system consists of two similar components whether

[1] For a fuller account of the aberrations of an optical system see "Optical Calculations," § (7), etc. See also "Microscope, Optics of the," § (5), etc.

of equal or unequal focal length. This is a simple case where the principle of symmetry applies, a principle of which great use is made in connection with other aberrations. It will be considered in more detail later.

It should be evident that little importance attaches to equality in the sizes of images of different colours when they occupy different planes. Accordingly, the chief attention should be given to the positions in which the images of various wave-lengths are formed, and if some latitude is necessary it may be given in a slight variation of magnification with the wave-length. Owing to the properties of the transparent materials available for the construction of lenses it will, as a rule, only be possible to secure exact agreement of focussing plane for two wave-lengths. The two regions it has been found most satisfactory to bring into agreement are the neighbourhood of the D lines of sodium and the G′ line of hydrogen, the former being not far from the dominant wave-length for visual observation, and the latter near the wave-length to which the photographic emulsions are most sensitive. This method of correcting enables a view to be correctly focussed for photographic recording by visual examination of the image.

(ii.) *Spherical Aberrations.*[1]—The remaining defects are of the type known to the optician as the spherical errors. The amount of these will also depend on the wave-length. They should usually be corrected for a wave-length in the neighbourhood of G′ rather than for D. If this is done the photographic record will present a better appearance than the visible image, and it may be a matter of importance to concentrate attention almost entirely on the portion of the image in the neighbourhood of the lens axis when judging the best position of focus visually. If the correction were made in the alternative way the visible image would appear very pleasing and the photographic record distinctly disappointing. In lenses for special purposes this consideration may not apply.

(*a*) *Distortion.*—In consequence of the spherical errors the image of a point in the object formed by light of a given wave-length will not on a ray theory be a point, or in one case its position, if a point, will be displaced from that which it should occupy were the image an exact projection of the object. This particular defect is known as distortion, and the displacement is necessarily directly towards or away from the point in which the image plane is met by the lens axis. The displacement may be represented mathematically by a series of odd powers of the distance the point would be from the axis were the defect absent. Thus, if y, z be the co-ordinates of a point in the image plane and we take the axis of y parallel to the direction in question, so that y represents that distance,

$$\delta y = a_1 y^3 + a_2 y^5 + a_3 y^7 + \ldots, \quad \delta z = 0.$$

The coefficients a_1, a_2, a_3 . . . may be used as measures of the extent to which the defect is present, and are named the coefficients of distortion of the first, second, third . . . orders. If the defect is serious it becomes obvious that lines in the image which correspond to straight lines in the object are appreciably curved. This want of straightness, as the above equation shows, gradually makes its appearance as the shortest distance between the line and the point of the image plane on the lens axis increases.

(*b*) *Spherical Aberration.*—Another aberration of a very simple kind is central spherical aberration, the "central" referring to the distinctive property that, though equally present in other parts of the field of view, it is only aberrations of this group which appear in images on the axis itself. This aberration consists essentially in the rays from narrow zones of the aperture, bounded by nearly equal circles centered on the axis, coming to foci on the axis which vary gradually from one zone to another. The rays which in an ideal lens should pass through a single point of the axis, in a lens suffering from this defect, will touch a caustic surface which has two branches, one being a surface of revolution about the axis and the other a length of the axis itself. If η, ζ are the co-ordinates of the point in which the ray meets the stop, the intersection of the ray with the image plane will be displaced from its ideal position given by y, z, by a distance whose rectangular co-ordinates δy, δz satisfy

$$\frac{\delta y}{\eta} = \frac{\delta z}{\zeta} = b_1 r^2 + b_2 r^4 + b_3 r^6 + \ldots$$

where $r^2 = \eta^2 + \zeta^2$. As before, b_1, b_2, b_3 . . . are aberration coefficients of the first, second, third . . . orders. In telescope objectives, the removal of this aberration is of outstanding importance, but it is common to find appreciable amounts of central spherical aberration present in good photographic lenses, the reason being that up to a certain point this defect is less harmful than some other aberrations which cannot in the particular design be removed simultaneously with this central aberration.

(*c*) *Curvature and Comatic Aberrations.*—Between the two simple aberrations considered in (*a*) and (*b*), there lie a series of others whose number and character depends upon the order of the aberration. Several of these have received names suggested by the shape of their trace in the image plane or by some other

[1] For a discussion of the expressions given below for the various aberrations see "Lens Systems, Aberrations of."

outstanding character. A very rough classification, based only on the aberrations of the first order, has been found sufficiently inclusive to group these under for testing purposes, and will serve for the present discussion. The loose character of the classification must, however, be borne in mind, as confusion may otherwise be caused when a more exact division is called for.

Under this grouping the remaining defects are classed as curvature errors or as comatic errors, the former class containing those members which tend to produce a symmetrical deviation from the ideal image point, the latter to those in which the departure tends to be unsymmetrical. Comparing this division with the two spherical aberrations already considered, central spherical aberration is evidently symmetrical in character, for corresponding to any displacement δy, δz in a ray from a given object point we obtain the complementary displacement $-\delta y$, $-\delta z$ by selecting the ray which meets the stop in $-\eta$, $-\zeta$ instead of η, ζ. On the other hand, distortion is an unsymmetrical aberration, since δy has a definite sign for all rays from a given object point. The characteristic feature of these aberrations in mathematical language is that the symmetrical aberrations may be represented by terms which contain odd functions of η and ζ, and even functions of y and z, while with the unsymmetrical aberrations the reverse is the case. Taking the first order aberrations, from which the classification is derived, it may be shown that the defect known as coma involves displacements of the ray intersections given by

$$\frac{\delta y}{cy}=2(\eta^2+\zeta^2)+(\eta^2-\zeta^2)$$

$$=r^2(2+\cos 2\theta),$$

$$\frac{\delta z}{cy}=2\eta\zeta=r^2\sin 2\theta,$$

where $\quad \eta=r\cos\theta,\quad \zeta=r\sin\theta,$

and c is a constant, showing that the rays which pass through a narrow circular zone of the aperture intersect the image in a circle, the radius of which is proportional to the square of the radius of the zone. The circle has the property that light from any one half of the annulus in the stop plane is distributed over the whole circle; it is thus formed twice over by the light from the complete annulus. The aberration is most likely to be detected in photographic prints by a fuzzy edge to one side of a narrow object, such as a pole, which stands out strongly against its background, while the other edge appears much sharper. The removal of this aberration is particularly important in photographic lenses on account of the large field of view involved, to which this particular aberration is proportional.

The remaining first order aberrations relate to departures of the image surface from the plane. Two independent constants e, e' occur, so that two aberrations are present, the formulae for the displacements being

$$\frac{2\delta y}{y^2}=\eta(3e+e')$$

and

$$\frac{2\delta z}{y^2}=\zeta(e+e').$$

It can be shown from these expressions that for points on a sphere of curvature proportional to $(3e+e')$ having its vertex in contact with the plane in which the image has been assumed to lie δy will assume zero value, and that similarly δz will vanish for points on a sphere whose curvature is proportional to $(e+e')$. These spheres are obviously the loci of the focal lines of the pencils which form the image. The aberration corresponding to e is called astigmatism, since, if e is finite, the two spheres are distinct, so that no point image exists. The other aberration denoted by e' is called the curvature, since this is the curvature the image surface would have if point images existed in the outer parts of the field of view. The form of the expression shows that it is only when e vanishes that the curvature can be represented by e', the possible range in other cases lying definitely to one side or other of this value, according to the likeness or unlikeness of the signs of e and e'.

The removal of aberrations of this type was for long an obstacle in the way of improvements of photographic lenses, and was not realised until the introduction of novel types of glass had resulted from the work of Abbe and Schott at Jena. These new glasses were of particular importance in enabling e' to be controlled in value. Together with this improvement it is necessary, if e is to receive a satisfactory value, for the optical system to contain lenses appreciably separated from one another.

(iii.) *Aberrations in General.*—In closing this very brief description of the aberrations to which lenses are subject, it is necessary to point out that the values of the constants by which the defects are measured are dependent on the position of the object. It follows that a lens which is corrected for an object in one position will not give so good an image when the object is considerably displaced. The magnitude of the defects depends largely on the size of the lens aperture, so that a lens of large aperture is necessarily less suitable for general work than one in which the aperture is smaller. In particular, if a lens corrected for an object at infinity is reversed and still used for a distant object the image will not be satisfactory unless the aperture is less than a definite value. An indication of this value is afforded by the largest aperture for which

symmetrical objectives are offered for the photography of distant scenes. The advantage to a manufacturer of the symmetrical construction is sufficiently great to ensure his utilising it as far as it can be satisfactory. A lens of this type cannot possibly be thoroughly corrected except for the reproduction of objects on full scale. Actually, such lenses are made for general work up to apertures of $f/5{\cdot}6$. This affords a direct means of finding the standard of accuracy which manufacturers find necessary in their objectives.

It must further be remembered that these first order effects are modified in the parts of the field of view more remote from the axis by the presence of aberrations of higher orders. These may tend in many instances to reduce the evil effects of the aberration of lower orders; for instance, the surfaces containing the focal lines may be caused to depart from the spherical form and made to approach the ideal image plane again. This condition is fulfilled in several modern objectives, the two focal surfaces crossing in or close to this plane near the edge of the useful field of view. Strictly speaking, this is the condition which is held to justify the application of the term "anastigmat" to a lens, though, on the one hand, this usage is far more free than mathematical conceptions warrant, and, on the other hand, the term is also applied by their makers to lenses which do not even attain this limited degree of stigmatism.

The actual appearance of the image in any lens depends upon the combined effect of all the aberrations which are present, and is also modified by the finite wave-length of the light forming the image. The classification of the defects into a number of separate aberrations is entirely artificial, its justification being the convenience of dealing with the defects in the manner described.

§ (9) Types of Lenses.—The most varied forms have been proposed and used for photographic lenses, and it is impossible here to attempt any detailed historical treatment of the subject. Much interesting matter will be found in the writings of von Rohr[1] and Gleichen,[2] which should be consulted by all who desire fuller information. A broad division of the forms in use at the present day may be drawn between the symmetrical and the unsymmetrical classes. The former group may be subdivided into

(*a*) Fully symmetrical;
(*b*) Hemi-symmetrical;
(*c*) Quasi-symmetrical;

and in each of these subdivisions there may be anastigmatic or astigmatic types, combinations built upon cemented or uncemented components, or on still more complex forms.

The unsymmetrical forms may be built from 1, 2, 3, 4, or 5 components, some or all of which may themselves be compound lenses. The cases of most importance are

(*a*) The single lens;
(*b*) The double lens;
(*c*) The telephoto lens, a special case of the double lens;
(*d*) The triple lens.

As practically all lenses of the symmetrical type are obtained by combining together two systems which are themselves unsymmetrical lenses of classes (*a*) or (*c*), it appears the simplest course to consider the unsymmetrical forms first.

§ (10) Unsymmetrical Types. (i.) *The Single Lens.*—This form of lens may be constructed from one or several pieces of glass cemented together. All lenses of this type are used with an aperture stop some distance in front of them, and are meniscus in external form, presenting their concave surfaces to the incident light. This arrangement is adopted to secure a pseudo-flattening of the image by selecting for different parts of the field of view pencils which have been refracted through different regions of the lens. The device is only effective when either central spherical aberration or coma or both are present in the image. The stop must not under these circumstances be large, and as this method of disguising curvature of the field necessarily introduces distortion, systems of this kind cannot be used on large fields containing numerous straight lines which do not lie near the lens axis. It is readily seen from the diagrams (*Figs.* 6 and 7) that the presence

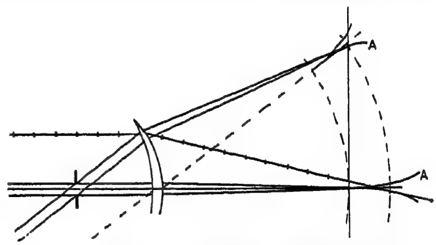


Fig. 6.—Curvature Correction in the Presence of Over-corrected Spherical Aberration.

of coma is of greater assistance than spherical aberration, and that when either is present alone the coma should be under-corrected and the spherical aberration over-corrected. The sign required for the outstanding coma accounts for the use of a lens strongly concave to the incident light.

In some of the more complex modern forms a ruling factor in the design is the presence of at least two cemented surfaces at which the angles of incidence may be large in comparison

[1] *Theorie und Geschichte des photographischen Objektivs.*
[2] *The Theory of Modern Optical Instruments*, trans. by Emsley and Swaine.

with those at other surfaces, one of these two being of positive power and the other of negative power, the system further containing

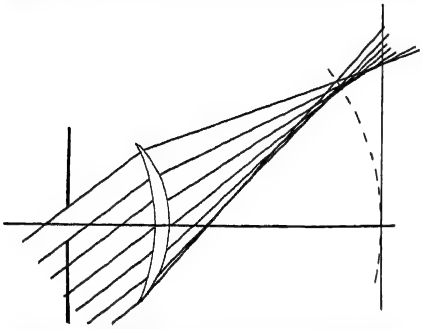


FIG. 7.—Curvature Correction in the Presence of Under-corrected Coma.

at least one lens of barium crown glass which combines a high refractive index with small dispersion. For a fuller account of such lenses reference may be made to the treatise by von Rohr, already mentioned, or to Lummer's *Photographic Optics*, translated by Thompson. The earliest lenses used in photography appear to have been single lenses of one glass of the form devised by Wollaston. The earliest achromatic lenses were of the ordinary telescope form with a double convex crown lens, but at the present time a form with both crown and flint of meniscus shape, and with the crown lens nearer the stop, is invariably used, as it yields more satisfactory images over an extended field.

(ii.) *Systems composed of Two Dissimilar Lenses.*—Such lenses when separated by an appreciable distance comprise one of the most important classes of objective. By far the most celebrated design which has yet been evolved, the Petzval portrait lens (*Fig.* 8),

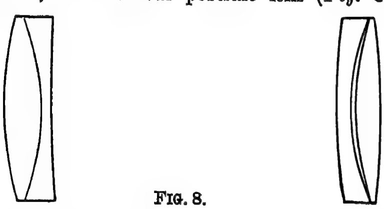


FIG. 8.

belongs to this group. This lens was produced in the very early days of photography, when the need for great rapidity was acute on account of the slowness of the sensitive materials at that time available, and in many respects it has hardly been surpassed by any later designs. In either its original form or in the modification introduced by Dallmeyer (*Fig.* 9) it is extensively used as a cinematograph projection lens at the present day, the relative aperture varying from $f/3{\cdot}5$ to about $f/2$. These lenses are, of course, made with the old silicate glasses.

The form of Petzval's objective generally corresponds to that which would be indicated by basing the design on the conditions for the

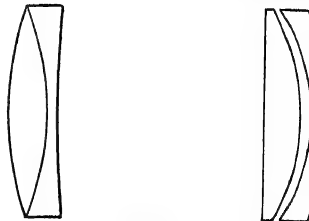


FIG. 9.

removal of first order aberrations as far as this is possible with the glasses available. As any actual design involves a balance between aberrations of different orders which cannot be effected simultaneously, the results given by the expressions for these first order aberrations are subject to considerable modification in the evolution of the final system.

The other members of this class which require special mention are anastigmatic lenses, and so utilise barium crown glasses. They are generally composed of an achromatic combination of the older types of glass, which is over-corrected for central spherical aberration, on one side of a central stop, and an achromatic combination of a light flint with a dense barium crown, which is necessarily under-corrected for central spherical aberration if the surfaces are spherical, on the other side of the stop. Both components are meniscus in external form, and present their concave sides to the stop, the ordinary achromat being of the form usual in single lenses but being placed before the stop, the other lenses resembling in shape the ordinary telescope objective with its flint component next to the stop. Under these conditions the cemented surface of the front component is of negative power and that of the back component of positive power. The need for presenting the concave sides of both components to the stop arises from the conditions for the removal of coma and distortion. Among the best known lenses of this class are the Protars of Messrs. Zeiss (*Fig.* 10). The Aldis lens, which in form is a double objective, is most simply regarded as a special case of a triple objective.

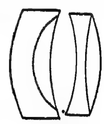


FIG. 10.

(iii.) *The Telephoto Objective.*—This form, as a self-corrected lens in which the components stand in a fixed relation to one another, is of comparatively recent development. The essential feature is that a positive lens is followed at a considerable distance by a

negative lens, which increases the focal length of the system without extending the separation between the front lens and the image in the same proportion. Such a combination of a positive and a negative lens, each corrected for colour, lends itself very readily to the construction of an anastigmatically corrected system Among these systems the large aperture telephoto lens (*Fig.* 11) made by Messrs. Taylor, Taylor & Hobson may be mentioned.


FIG. 11.

Although telephoto lenses made of two single glasses are not employed as an unsymmetrical objective, such systems require mention, as they provide the basis of many symmetrical objectives. In particular, the Gaussian form of telescope objective is used in more than one of these more complex lenses.

(iv.) *Lenses with Three Components.*—Among the early attempts to improve photographic objectives systems consisting of three separated components are of frequent occurrence, the external components being of positive power and the central one of negative power. In most instances each component was achromatised independently. None of these lenses survive to-day, and their present interest is chiefly due to the claim they may have to be considered as forerunners of the Cooke lens (*Fig.* 12), designed by Mr. H. Dennis Taylor of Messrs. T. Cooke & Sons, and made by Messrs. Taylor, Taylor & Hobson. When account is taken of the large number of conditions which have to be satisfied and the simplicity of the means adopted to achieve them, the design is one which must command general admiration. The excellence of its actual performance is sufficiently attested by the reputation of the lens among photographers. The general form of the lens, as in other cases, may be derived by considering aberrations of the first order only.

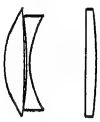


FIG. 12.

Another well-known lens of this type which may be considered a slight elaboration of the Cooke is the Tessar (*Fig.* 13), produced by Messrs. Zeiss. The chief difference between the two lenses is that in the Tessar the back lens of the Cooke is replaced by a cemented doublet. In theory at least the additional degrees of freedom attained by the introduction of one more glass should enable a somewhat larger field of view to be covered with satisfactory definition by the Tessar than by the Cooke: whether in fact the gain, if any, actually achieved, justifies the complication, is debatable. Some indication of the excellent results attainable with the Cooke design is provided by the extensive use of such lenses of large aperture and long focal length by the German army in the war for aeroplane photography. So far as is known to the writer, the Tessar was not used for this purpose. The only serious drawback to lenses of this type appears to be the rapid deterioration in the definition outside the field which the lens is intended to cover, so that the extent to which the lens can be decentered is distinctly small compared with that possible in some of the double symmetrical types when a small stop is used.

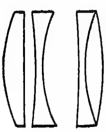


FIG. 13.

The Aldis lens (*Fig.* 14) may be regarded as a modified Cooke, in which the first two lenses are of substantial thickness and are cemented together. The reduction in the number of glass air surfaces which results from this modification is of distinct advantage, but the accompanying restriction in the number of variables at the disposal of the designer greatly increases the difficulty of attaining as high a standard of definition as with a true triple lens. These difficulties have been ably met, and this lens appears to be as close an approach as is possible with spherical surfaces to the highest triple standard.

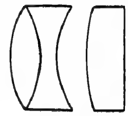


FIG. 14.

Many miscellaneous types of unsymmetrical lenses have been constructed by combining on the different sides of a central stop dissimilar types of double and triple lenses.

§ (11) SYMMETRICAL LENSES. (i.) *True Symmetricals.*—Turning now to lenses of the symmetrical class, the true symmetricals, obtained by combining together two exactly similar lenses of equal focal length in such a manner that one occupies the position of the image of the other in a plane mirror coincident with the central stop plane, may be considered first. Among the simplest of these lenses is one due to Sutton, in which water is the refracting agent, this being enclosed between two thin concentric glass hemispherical shells. The use of such shells enclosing various liquids has been repeatedly suggested, but, though satisfactory pictures have been produced by their means, they have not come into use for obvious reasons. Closely allied to this lens is the Hypergon of Messrs. Goerz (*Fig.* 15), consisting of two very deep positive meniscus lenses of a single glass, so placed that the external surfaces very approximately form a sphere. This lens is used

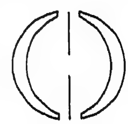


FIG. 15.

with a rotating star stop to equalise the illumination over the plate, and covers an angular field several times as great as is attempted with lenses of other types. For a long period the best lens for general purposes was obtained by using two achromatic meniscus lenses of the type described earlier to form a symmetrical lens. Many names were in use for such lenses (*Fig.* 16), though in essentials they differed very little from one another. The names by which they became best known are the "rapid rectilinear" and the "aplanat," the former describing the characteristic improvement on the single lenses used previously, and the latter description being justified by the improved definition which enabled a larger relative aperture to be employed, though in the modern sense in which aplanatic is used the term is not correct.[1]

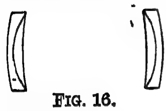


FIG. 16.

Among modern constructions of the anastigmatic types many are of this symmetrical class. The simplest elements from which they can be constructed in the case of cemented components must contain at least three different glasses to yield cemented surfaces whose powers differ in sign. The commonest types have either (*Fig.* 17) a positive

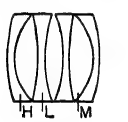


FIG. 17.

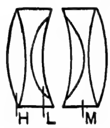


FIG. 18.

H = Highest index.
M = Medium index.
L = Lowest index.

double convex lens cemented between two negative lenses, or (*Fig.* 18) a double concave lens cemented between two positive lenses. In either case the refractive indexes of the glasses form a descending sequence on proceeding from the external medium towards the stop. In another form (*Fig.* 19) a positive meniscus lens

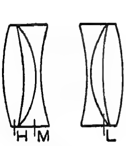


FIG. 19.

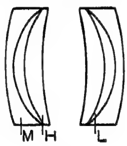


FIG. 20.

of low refractive index is placed between a double convex lens of a dense barium crown and a double concave lens of a light flint, the latter of course being nearest the stop. In a fourth form (*Fig.* 20) the central meniscus lens presents its concave instead of its convex side to the stop, and all the component lenses are meniscus, the negative lenses occupying the extreme positions. Still more complex forms are found if more than three glasses are used, and lenses of this class having four and five glasses to a single component are made.

In another class of symmetrical objective the central lens of the third type of triple cemented component is replaced by a meniscus air lens (*Fig.* 21), each component thus consisting of a double convex dense barium crown lens separated from a double concave lens made from a light flint glass. In a limiting case this type merges into the triple lens of the Cooke class. One further class in which

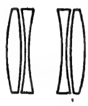


FIG. 21.

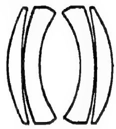


FIG. 22.

an air-gap is used remains, corresponding to a form of triple cemented component which does not appear to have been utilised. In this class the objective may be considered a derivative of the Gaussian objective, consisting of separated positive and negative meniscus lenses. The best known lens of this type is the Homocentric of Messrs. Ross (*Fig.* 22). A more complex lens constructed for reproductive work and available with apochromatic correction is the Planar of Messrs. Zeiss (*Fig.* 23). The cemented lens in

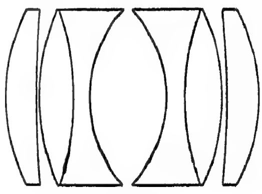


FIG. 23.

this construction is composed of two glasses with different dispersions but approximately equal refractive index.

(ii.) *Theoretical Considerations.*—It has often been claimed for the symmetrical construction that its use automatically removes distortion, coma, and all aberrations of the unsymmetrical class, whatever the position of the object may be. As this question is of very great importance it is worth inquiring what advantages a symmetrical construction does necessarily secure. For this purpose it is quite unnecessary to consider the details of the lens

[1] In earlier writings aplanatic signifies merely freedom from central spherical aberration. Abbe introduced the meaning now generally adopted by imposing the additional condition that the sine condition must be satisfied, so that the term now implies simultaneous correction for central spherical aberration and for coma.

construction, the symmetry itself being the only pertinent factor.

Suppose, then, that C (*Fig.* 24) is the centre of the stop, and thus symmetrically situated with respect to the two components. Any ray through C will strike the components in corresponding points and will be refracted by the separate components into exactly corresponding positions. The incident and emergent rays will thus strike similar and symmetrically situated surfaces in points which also correspond. In particular, if the intersection with the object and image planes for magnification −1 are P and P′, the distances of these points from the axis are equal,

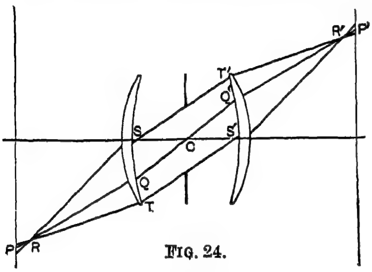


FIG. 24.

and the objective is therefore free from distortion for this particular magnification.

Again consider two parallel rays ST′ and TS′ in the space separating the components, which meet the central stop plane in diametrically opposite points. Symmetry again shows that the incident ray corresponding to ST′ is parallel to and correspondingly situated with the emergent ray belonging to TS′, and *vice versa*. Thus the intersection R of the two incident rays is symmetrically situated with respect to the intersection R′ of the emergent rays. Moreover, the previous result shows that the ray from R through C passes through R′, and the more general case is easily established, that corresponding to any ray through R and R′ another ray through the same two points may be found which is parallel to the former ray in the stop space and meets the stop in the diametrically opposite point. These results are readily interpreted to mean that the lens symmetry automatically secures freedom from all unsymmetrical aberrations when the magnification is −1. It is to be observed that there is no need for the various pairs of rays such as ST′ and TS′ to be parallel in the stop space to QQ′ or to other pairs of rays arising from the same object point, so that the uncorrected character of the half lens is of no consequence. When we proceed to consider conjugate planes for a different magnification the arguments which have served to establish this result for the magnification −1 break down. For example, in the case of distortion, freedom from this aberration evidently involves the absence of aberration in the image of C in a single component. To secure the absence of all chromatic aberration the single component must bring a parallel bundle of rays incident in the stop space to a single focus in the image space, and must also form an image of C in the same place for the different colours.

When coma is considered in a similar way for a magnification other than −1 the rays concerned in the formation of the image of a particular object point which intersect the stop in opposite ends of a diameter do not lie parallel to one another in the stop space, and in the absence of such symmetry conclusions can only be drawn when more is known of the unsymmetrical aberrations for the magnification −1. Algebraic investigation is in this case more simple, and leads definitely to the conclusion that with the symmetrical construction the only magnifications for which the aberrations generally can be removed are +1 and −1, the former corresponding to the stop and the latter to a real image.

With moderate sized stops the fact that a symmetrical objective is free from unsymmetrical aberration when copying full size may be taken to involve small but not zero values for the unsymmetrical aberrations for objects in other positions, but for large apertures the conclusion is unavoidable that this construction is undesirable. The design of a symmetrical objective for copying the same or approximately the same size is particularly simple, since it reduces to the design of a half component which is free from symmetrical aberrations alone for parallel beams incident on the surface presented to the stop, together with the correction of the image of the point in which the stop meets the axis for colour and spherical aberration. The conditions to be satisfied are thus considerably fewer in number than with a complete unsymmetrical objective.

(iii.) *Hemi-symmetrical Objectives.*—These are formed by combining together two components of similar design but of different focal lengths. The position of the stop should divide the space between the inner surfaces in proportion to the focal lengths of the components, and, as in the fully symmetrical systems, the image of the axial point of the stop plane should be free from chromatic and spherical aberrations. The want of symmetry which exists in the paths of other rays shows that we cannot in this case infer absence of unsymmetrical aberrations for the magnification determined by the ratio of the focal lengths of the components, though when these focal lengths are not extremely dissimilar in magnitude considerations of continuity lead us to expect that aberrations of this kind will not be large. Lenses of this type are chiefly used as a variation of the fully symmetrical type to secure a lens of somewhat different focal length at the cost of half a lens, one component only of the fully symmetrical lens being replaced by another of the same design but of different focal length.

(iv.) *Quasi-symmetrical Systems.*—It is evident from the preceding discussion that for the best results for distant objects with large apertures the symmetrical construction is unsuitable. In a number of lenses this difficulty is overcome by introducing slight differences between the two components, as by altering one or two curvatures and possibly a separation also in one component compared with the other. In this way the advantages which the symmetrical construction offers to a manufacturer may in part be retained,

and the design of the lens is also simplified, since each component must be given properties which differ in known directions from those which yield correct images for full-size copying.

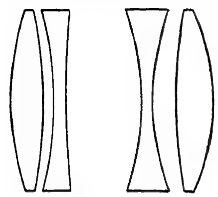


FIG. 25.

Such lenses may be called quasi-symmetrical objectives, and among them may be noted the Celor (*Fig.* 25) made by Messrs. Goerz with an aperture of $f/3{\cdot}5$, while several forms of the Cooke lens and of its variants may be regarded as special cases of quasi-symmetrical lenses. In most of these objectives the outstanding feature of the modification is that the front portion, in comparison with the back, has surfaces more strongly convex to light proceeding from external space to the stop.

T. S.

PHOTOGRAPHIC LENSES FOR TELESCOPES. See "Telescope," § (9).

PHOTOGRAPHIC METHOD OF SPECTROPHOTOMETRY: a method of great value in the ultra-violet part of the spectrum. See "Spectrophotometry," § (17).

PHOTOGRAPHIC REPRODUCTION OF GRATICULES. See "Graticules."

PHOTOGRAPHIC SURVEYING. See "Surveying and Surveying Instruments," § (9) (vii.).

PHOTOMETRIC METHODS OF TEMPERATURE DETERMINATION. See "Photometry and Illumination," § (112).

PHOTOMETRY AND ILLUMINATION

I. GENERAL PRINCIPLES AND DEFINITIONS

§ (1) GENERAL.—Photometry is that branch of applied physics which deals with the quantitative comparison of light sources and their evaluation in terms of some agreed standard. It will be seen from Part III. on "Photometric Methods" that this comparison or evaluation is always achieved by means of a comparison of the brightnesses of surfaces illuminated by the sources, and this comparison has always to be made by the human eye. There is therefore a natural limit to the accuracy with which two sources of light can be compared, viz. the limit beyond which the human eye is incapable of perceiving differences of brightness. This limit varies with the absolute value of the brightnesses compared, any difference that may exist between their spectral distribution, the state of the eye, and other factors which are discussed elsewhere (see article on "The Eye as an Optical Instrument").

In photometry there is always an ordered sequence of phenomena each of which has to be separately considered. Some object, called a "luminous source," emits radiation which is capable of affecting the retina of the human eye and thereby producing the sensation of light. This radiation, or some portion of it, reaches a given surface and thereby "illuminates" it; the surface reflects some portion of the radiation it receives and then is said to have a certain "brightness."

The consideration of this sequence leads to a system of definitions of photometric terms. These may be based on the luminous power of the source as the primary conception,[1] or alternatively, the rate of flow of radiation (*i.e.* luminous flux) may be regarded as fundamental, all other phenomena, including the source, being referred to this as the primary conception.[2]

§ (2) DEFINITIONS.—The following definitions of the more important photometric quantities may be given. Those marked with an asterisk (*) were adopted by the International Commission on Illumination in 1921.

*(i.) *Luminous Flux.*—It is the rate of passage of radiant energy evaluated by reference to the luminous sensation produced by it.

Although luminous flux should be regarded strictly as the rate of passage of radiant energy, as just defined, it can, nevertheless, be accepted as an entity, for the purposes of practical photometry, since the velocity may be regarded as being constant under those conditions.

It is measured by the quantity of luminous energy falling per second on a surface of unit area placed at right angles to the flow.

*(ii.) *The Unit of Luminous Flux* is the Lumen. It is equal to the flux emitted in a unit solid angle by a uniform point source of one international candle.

(iii.) *A Luminous Source*, or source of light, is one which emits luminous flux.

*(iv.) *The Luminous Intensity* (*candle-power*) of a point source in any direction is the luminous flux per unit solid angle emitted by that source in that direction. (The flux emanating from a source whose dimensions are negligible in comparison with the distance from which it is observed may be considered as coming from a point.)

*(v.) *The Unit of Luminous Intensity* (*candle-power*) is the international candle, such as resulted from agreements effected between the three National Standardising Laboratories of France, Great Britain, and the United States, in 1909.

[1] A. P. Trotter, *Illum. Eng.*, London, 1914, vii. 339.

[2] "Report of the Committee on Nomenclature and Standards of the Illuminating Engineering Society," *Am. Illum. Eng. Soc. Trans.*, 1917, xii. 438.

This unit has been maintained since then by means of incandescent electric lamps in these laboratories which continue to be entrusted with its maintenance.

These laboratories are: the Laboratoire Central d'Électricité in Paris, the National Physical Laboratory in Teddington, and the Bureau of Standards in Washington.

The International Candle is thus an arbitrary standard in terms of which luminous intensity is measured.

* (vi.) *The Illumination* at a point of a surface is the surface-density of the luminous flux at that point or the quotient of the flux by the area of the surface, when the latter is uniformly illuminated.

* (vii.) *The Practical Unit of Illumination* is the Lux. It is the illumination of a surface one square metre in area receiving a uniformly distributed flux of one lumen, or the illumination produced at the surface of a sphere having a radius of one metre by a uniform point source of one international candle situated at its centre.

In view of certain recognised usages, illumination may also be expressed in terms of the following units:

Taking the centimetre as the unit of length, the unit of illumination is the lumen per square centimetre; it is known as the Phot. Taking the foot as the unit of length, the unit of illumination is the lumen per square foot; it is known as "Foot-candle."

1 "foot-candle" = 10·764 lux
1·0764 milliphot.

(viii.) *The Brightness* of a luminous surface in a given direction is the candle-power per unit projected area of the surface in that direction. It is expressed in candles per square millimetre or per square metre.[1]

N.B.—Confusion has been caused in the past by the use of the illumination unit "foot-candle" or its modification "equivalent foot-candle" to express brightness. Brightness should preferably be expressed in terms of candles per unit area.

(ix.) *The Mean Horizontal Candle-power* of a lamp is the average candle-power in the horizontal plane passing through the luminous centre of the lamp.

It is here assumed that the lamp (or other light source) is mounted in the usual manner, or, as in the case of an incandescent lamp, with its axis of symmetry vertical.

(x.) *The Average Candle-power* of a lamp is the average value of the candle-power measured in all directions in space. It is numerically equal to the total luminous flux emitted by the lamp in lumens divided by 4π. This has been generally termed the "mean spherical candle-power," and similarly the term "mean hemispherical candle-power" of a lamp (upper or lower) has been used for the average candle-power of a lamp in the hemisphere considered. It is recommended that for this purpose the name of the hemisphere in question should be added to the term "average candle-power," thus: "average candle-power (upper hemisphere)." In the past the term "Efficiency" has been generally used to denote the performance of a lamp in terms of the watts of electrical power consumed, divided by the luminous intensity in candles. It is recommended that this should be called the "watts per candle" of the lamp.

[1] Normal brightness is the brightness of a surface when viewed in the direction of its normal.

§ (3) Fundamental Relation of Photometry.—From the fact of the rectilinear propagation of luminous radiation it at once follows that, with a uniform point source: (*a*) the flux incident per unit area, on a very small surface normal to the direction of the light, varies as the inverse square of the distance of the surface from the source; and (*b*) the flux incident per unit area, on a very small surface at a given distance from the source, varies as the cosine of the angle between the normal to the surface and the direction of the light. Thus, if the distance between a surface and a luminous source so greatly exceeds the dimensions of both that these may be neglected by comparison, the illumination E of the surface varies (i.) directly as the candle-power J of the source in the direction of the surface; (ii.) inversely as the square of the distance of separation, d; (iii.) directly as the cosine of

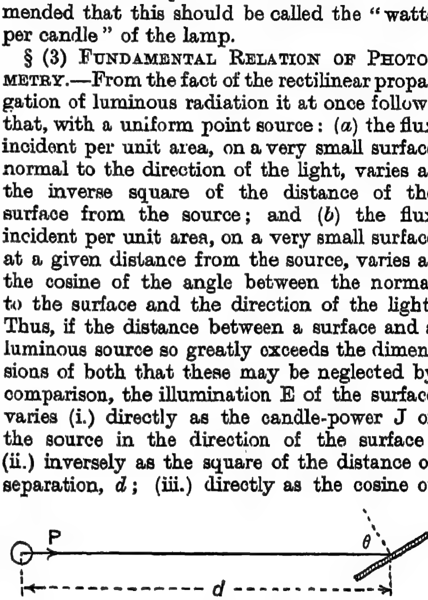


Fig. 1.

the angle θ between the normal to the surface and the line joining it to the source (see *Fig.* 1), or, in symbols,

$$E = J \cos \theta / d^2.$$

The second and third of these laws are known as the "inverse square law" and the "cosine law" of illumination respectively, and it is upon the equation just given that the whole of the science and practice of photometry is based.

II. Photometric Standards

The standards to be dealt with in this article include two groups: (*a*) primary standards, in terms of which the candle-powers of all sources are expressed; and (*b*) secondary standards, for practical use in everyday photometry.

§ (4) Requirements of Primary Standards.—The conditions which a primary standard should fulfil are those required of any physical standard, viz. ease of reproducibility

from specification, maintenance of value over long periods, small correction factors for change of conditions such as barometric pressure, temperature, etc. In addition to these, a standard of candle-power must fulfil, as far as possible, the condition that the spectral distribution of its light shall approximate to that of the light sources measured by comparison with it. This condition is due to the great difficulties introduced into photometry by difference of colour between the lights being compared (see § (102) *et seq.*).

Many suggestions have been made at different times for the production of a satisfactory primary standard of candle-power. None of these fulfils, even approximately, all of the conditions outlined above. Most are difficult to reproduce with sufficient accuracy and are greatly affected by change of exterior conditions. None fulfils the condition as to spectral distribution, the colour of the light being, in every case, much redder than that of the sources of light in common use to-day.

§ (5) Kinds of Standards.—The proposed standards which have so far received any degree of support may be divided into two groups: (i.) flame standards, depending on the candle-power of a luminous flame burning under specified conditions; (ii.) incandescence standards, depending on the candle-power of a specified area of some solid or molten material at a given temperature.

Of the flame standards the most important are the British standard candle, the Vernon-Harcourt Pentane Lamp, and the Hefner Amyl-Acetate Lamp.

§ (6) British Standard.—(i.) The British candle was set up under the Metropolitan Gas Act of 1860 as the official standard for the purpose of testing London gas. It was described as a spermaceti candle weighing one-sixth of a pound, and burning at the rate of 120 grains per hour.

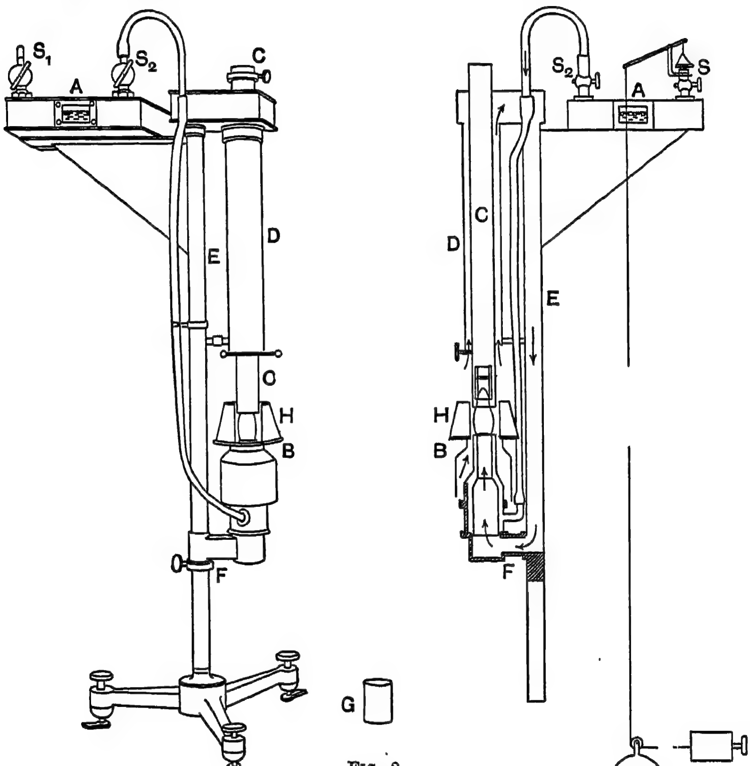


Fig. 2.

(ii.) This standard has now been superseded for practical purposes by the pentane lamp devised by Vernon-Harcourt in 1877. This lamp is shown diagrammatically in *Fig.* 2. The saturator A holds the fuel, liquid pentane, a highly inflammable and very volatile hydrocarbon distilled from petroleum. This saturator is filled to about two-thirds of its capacity before the lamp is lighted. The level of the liquid (observed through the window in the side of the saturator) is never

allowed to fall below one-eighth of an inch when the lamp is in use. The saturator is connected, by means of a wide india-rubber tube, with the burner B, which consists of a steatite ring pierced with 30 holes, and at which the issuing mixture of air and pentane vapour is ignited. The rate of flow of the mixed vapour, and therefore the height of the flame, can be adjusted by means of the stop-cocks S_1 and S_2 on the saturator. The chimney tube CC is furnished near its base with a mica window, upon which is marked a horizontal line 88 mm. above the bottom of the chimney. The chimney is set, by means of a cylindrical wooden gauge G, so that its lower end is exactly 47 mm. above the steatite ring burner. Surrounding the chimney CC is a concentric tube D up which a current of air is drawn by the heating of the chimney, and this heated air passes into the hollow supporting pillar E and so down through F to the centre of the steatite ring where it is used in the combustion of the pentane. It is important that the chimney CC be brought centrally over the burner B, and three screws are provided at the base for the purpose of making this adjustment. H is a conical shade for protecting the flame from draught. When in use, the lamp is set up with the pillar E vertical, and the stop-cocks S_1 and S_2 are so adjusted that the tip of the flame just rises to a level halfway between the bottom of the mica window in CC and the cross bar. (A slight variation in the height of the flame, however, does not affect its candle-power.) The mica window must, of course, be turned away from the photometer head, while H is turned so that the whole of the flame is visible from the photometer, except the portion at the top which is cut off by the lower part of the chimney. The saturator A is at first placed upon its bracket as far from the central column as possible and the lamp is left alight for at least a quarter of an hour before any photometric measurements are made. If it is found at the end of this period that the flame has a tendency to fall in height, the saturator is moved slightly towards the central column. In making photometric measurements all distances are reckoned from the centre of the flame, *i.e.* the geometric centre of the steatite ring.

(iii.) *The Pentane Lamp as Standard.*—When burnt under standard conditions of temperature, pressure, and humidity, the pentane lamp is recognised as having a candle-power of ten international candles. For a more accurate and detailed specification of the lamp, and of the preparation of the pentane, the Notification of the Metropolitan Gas Referees for the year 1916, published by H.M. Stationery Office, should be consulted. A copy of this notification, together with working drawings of the pentane lamp, is deposited at the National Physical Laboratory, and may be there consulted.

The value of the candle-power of the pentane lamp depends on the humidity and barometric pressure of the atmosphere in which it is burning. Several determinations have been made of the effect of these variables on the candle-power of the lamp. The latest is that of C. C. Paterson and B. P. Dudding at the National Physical Laboratory,[1] and for the details of their experiments their original work should be consulted. The result they obtain is given by the formula

$$\text{C.P.} = 10\{1 + 0{\cdot}0063(8 - e) - 0{\cdot}00085(760 - b)\},$$

where C.P. represents the candle-power of the lamp when burning in an atmosphere at a pressure of b mm. of mercury with a humidity of e litres of water-vapour per cubic metre of the moist air. The constants of the above equation agree exceedingly well with those given by Butterfield, Haldane, and Trotter,[2] but Rosa and Crittenden [3] find values for the constants which are notably lower than those given above, viz. 0·00567 and 0·0006. It has been suggested that this difference in the constants may be accounted for on the assumption that the lamp really possesses a temperature coefficient, but that as humidity and temperature are so closely related in any one locality, their separate effects cannot be determined by the usual method of observation. Differences in the relation between these two quantities at Washington and at Teddington might explain the observed differences in the humidity coefficients found at these two places. The values of the constants are obtained from a very large number of comparisons with an electric glow-lamp sub-standard of the same colour, observations being made under all available conditions of pressure and humidity. The most probable values of the constants are then found by the method of least squares. The value for the humidity correction found in 1917 by K. Takatsu and M. Tanaka [4] was 0·00638, and they suggest that the different value found in America may be due to the use of a hood and ventilating duct, if the effect of this is not the same at all humidities.

§ (7) The Hefner Lamp.—The standard of candle-power adopted as legal in Germany and some other European countries is the lamp devised in 1884 by von Hefner Alteneck and shown in *Fig.* 3. It consists of a container C, made of brass, 70 mm. in diameter and 38 mm. high. It holds about 115 c.c.

[1] "The Unit of Candle Power in White Light," *Phys. Soc. Proc.*, 1915, xxvii. 281.

[2] *Journ. of Gas-Lighting*, 1911, cxv. 88.

[3] *Am. Illum. Eng. Soc. Trans.*, 1910, v. 753.

[4] Electro-Techn. Laborat., Dept. of Communications, Tokyo, Oct. 1917.

of amyl acetate, a specially pure grade of this compound being required for photometric purposes. The liquid should always be

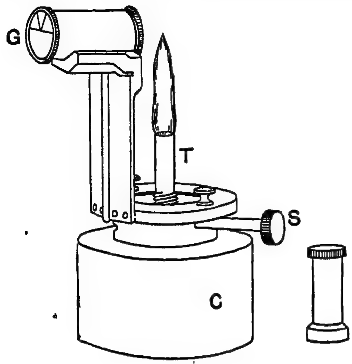


Fig. 3.

emptied out of the container when the lamp is not in use, as otherwise corrosion is liable to take place even though the inside of the container be tinned or nickel-plated. A thin German silver tube T, constructed very accurately to the dimensions of 25 mm. in height, 8 mm. in internal diameter, and 0·15 mm. in thickness of metal, holds a wick of 15 to 20 strands of untwisted cotton which can be adjusted in height by means of the screw S. G is a gauge consisting of a lens and ground glass screen with a horizontal cross-line. The lens forms an inverted image of the flame on the screen, and by this means the tip of the flame can be very accurately adjusted to the correct height of 40 mm. above the level of the tube. The candle-power of the flame depends appreciably on its height, and one of the chief disadvantages of this standard is the fact that the flame is very lambent and sensitive to draught, so that in the absence of any chimney its use is attended with great difficulty for practical measurement.

Liebenthal finds that a variation of 1 mm. in the height of the flame causes 2·7 per cent change in candle-power. The gauge on the lamp may be raised 4 mm. to permit adjustment of the flame to a height of 44 mm., at which height the candle-power has been found to approximate very closely to the unit derived from the pentane lamp.

With the Hefner, as with the pentane lamp, allowance has to be made for pressure and humidity of the air. The formula in the case of the Hefner lamp is

$$\text{C.P.} = 1 + 0{\cdot}006(8{\cdot}8 - e) - 0{\cdot}00011(760 - b),$$

where e and b have the same meaning as before.[1]

§ (8) The Carcel Lamp.—The flame standard officially adopted in France for gas-testing is the Carcel lamp, which is a lamp burning colza oil with an Argand burner and glass chimney. The wick is continually supplied with oil by means of a clockwork pump, and the correct candle-power is obtained when the rate of oil consumption is 42 gm. per hour. In actual use the wick is adjusted to give approximately this consumption, and a correction is applied for the departure from this theoretical value. The actual consumption is obtained by burning the lamp on a form of balance, and noting the time at which the lamp loses ten grammes in weight. The difficulties attending the use of this lamp are very great, and different observers are unable to obtain results either consistent among themselves or in agreement with one another.

§ (9) The Violle Standard.—Of the incandescence standards which have been proposed two require some description, viz. the Violle platinum standard, and the positive crater of the carbon arc. The former, which was proposed in 1881 by Violle, is the light from one square centimetre of a surface of molten platinum at the temperature of solidification. The Violle standard, on account of its obvious theoretical advantages, has been regarded as a promising advance on the existing flame standards, and unsuccessful attempts have been made at various times to place it on a satisfactory practical basis. The most recent careful work on this standard is that of Petavel,[2] who used a semicircular bar of platinum heated by an electric current to such a temperature that the inner core of the bar melted while the outer shell remained solid. The second form of the standard on which he made measurements was an ingot of platinum fused in a crucible of pure lime by means of an oxy-hydrogen blow-pipe. The metal was first completely melted and then heating was stopped and photometric measurements of the brightness were made at 10-second intervals during cooling. The readings when plotted showed a constant value over the region corresponding to the time of solidification, and the mean of the observations at this period was taken as the value required. It was found that the values obtained by this method did not depend on the shape or mass of the ingot, but that the effect of contaminating the platinum with either silica or carbon was very marked. Petavel's final conclusion was that the probable variation in the light emitted by molten platinum under standard conditions was not greater than one per cent, and that with more experimental refinements an even greater accuracy than this might be attainable. It cannot be said, however, that one per cent is sufficient to bring the Violle standard, with its additional disadvantage of redness of light,

[1] E. Ott, *Journ. of Gas-Lighting*, 1915, cxxxii. 378; and *Electrician*, 1915, lxxvi. 227.

[2] *Roy. Soc. Proc.*, 1899, lxv. 481.

into serious competition with the existing standards.

§ (10) LUMMER-KURLBAUM STANDARD.—One other platinum standard which has been proposed, and may be mentioned in passing, is that of Lummer and Kurlbaum. This consists of the light emitted by one sq. cm. of a platinum strip raised to such a temperature that a layer of water 2 cm. in thickness transmits 10 per cent of the total resultant radiation. This ratio is determined by means of a bolometer. Although this standard is used at the Physikalisch-Technische Reichsanstalt for the checking of Hefner lamps, Petavel has found (*loc. cit.*) that the bolometer method of temperature adjustment is not sufficiently exact to enable this apparatus to fulfil the conditions of a primary standard.

§ (11) CARBON ARC STANDARD.—The other incandescence standard, on which a considerable amount of work has been done by Morris, Forrest,[1] and others, is that provided by a square millimetre of the positive crater of a carbon arc operating under conditions designed to ensure steadiness. In the Forrest arc (*Fig.* 4) two negatives are employed, each at an angle of about 100° with the positive. Under these conditions, and using carbons of 8 mm. diameter with a total current of 7 to 10 amperes, it was found that the brightness of the crater was uniform over the whole of its surface, and photometric measurements were made of the candle-power per sq. mm. by inserting in front of the crater at D a small diaphragm of accurately known dimensions. Forrest found that the arc would work quite silently over a considerable range of currents, and that the crater brightness was independent of the current under these conditions. The value he obtained was 172-174 candles per sq. mm., and more recent work on the same lines by Allen[2] has confirmed his general conclusions and gives 176 candles per sq. mm. as the intrinsic brightness of the positive crater of a silent arc.

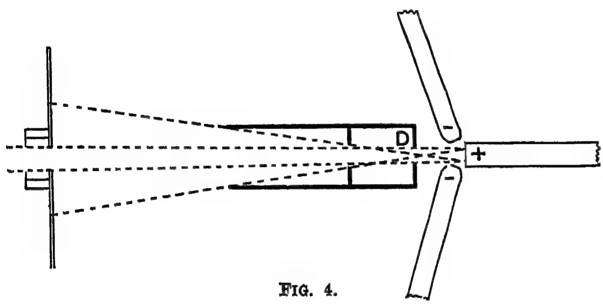


FIG. 4.

An advantage which this standard possesses over any other is that the colour of the light is bluer than that of most present-day sources, a difference which will diminish as the efficiency of practical illuminants increases.

§ (12) BLACK BODY STANDARD.—The proposal to use a definite area of a total radiator (black body) at some definite temperature has been made during recent years,[3] and as early as 1908 Waidner and Burgess suggested the use of one sq. centimetre of a black body at the temperature of melting platinum. The chief difficulty is that, as with all incandescence standards, the light emitted varies as a high power of the temperature of the radiator, so that the accuracy of the temperature measurements needs to be at least ten times as great as the desired accuracy of the light standard.

§ (13) GLOW-LAMP SUB-STANDARDS.—Finally, it is necessary to describe the electric glow-lamp sub-standards which are used for daily work in practical photometry. Specimens of such lamps as used at the National Physical Laboratory are shown in *Fig.* 5, some having carbon and some drawn tungsten filaments. In the case of the latter the filaments are disposed in grid formation so that the plane of the whole filament can be brought accurately over any mark on the photometer bench. Such standards as these can be operated at any given efficiency so that an approximate colour match may be obtained with the source with which they are to be used. Paterson has found[4] that, if not overrun, these lamps will maintain their

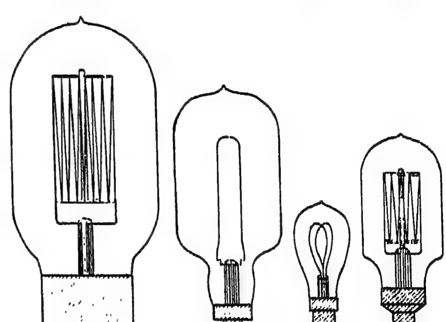


FIG. 5.

[1] *Electrician*, 1913, lxxi. 720 and 1007.
[2] *Phys. Soc. Proc.*, 1921, xxxiii. 62.
[3] E. Warburg, *Zeits. Vereines Deutsch. Ing.*, 1917, xxx. 3-10.
[4] *Inst. El. Eng. Journ.*, 1907, xxxviii. 287, and *Phys. Soc. Proc.*, 1915, xxvii. 270.

candle-power values for many hours of burning, so that if used only for a few minutes each day, they do not need recalibration for long periods. For this reason they have very largely superseded the actual primary standards in everyday photometry, and standards such as these, after comparison with master standards at a National Laboratory, are used in most photometric laboratories for practical work. Since the candle-power of a carbon filament varies as the sixth and that of a tungsten filament as the fourth power of the voltage applied to it, it is necessary to make the electrical measurements with 4 to 6 times the accuracy required of the candle-power measurements. With proper battery supply and potential measuring apparatus, however, this presents little difficulty.

§ (14) Relative Values of the Standards.—The following table summarises the relative values of the different standards above described.

Table I

	Pentane.	Hefner.	Carcel.	Violle.
Pentane .	1	11·1	1·05	0·5
Hefner . .	0·09	1	0·094	0·045
Carcel . .	0·96	10·6	1	0·48
Violle . .	2·0	22	2·1	1

Since 1909 the standard derived from the 10-candle pentane lamp has been adopted by the National Laboratories of France, Great Britain, and U.S.A., and has been subsequently preserved by means of glow-lamp sub-standards. This is now termed the International Candle, and is equal to the old French "Bougie décimale."

III. Photometric Methods

As has already been stated in the section on General Photometric Principles (*q.v.*), the comparison of light sources is always made by comparing the brightness of surfaces illuminated by these sources. Thus every photometer is essentially an instrument for the ready comparison of brightness. Photometers may, however, be grouped in two classes according as they are primarily intended for the comparison of light sources, or for the measurement of illumination at a given position.

In the case of the former class the two sources are generally placed one on either side of the photometer, and their respective distances from the comparison surfaces in the instrument are then adjusted until equality of brightness is obtained. Assuming both the surfaces to have the same reflection ratio, and to be equally inclined to the incident light, this condition gives the distances at which the two sources produce equal illuminations, and hence the candle-powers of the sources (in the direction of the photometer) are in the same ratio as the squares of their respective distances from the photometer surfaces.

§ (15) The Bunsen Photometer.—As a convenient example, the simple form of the Bunsen Grease-Spot Photometer may be described. This consists of a sheet of opaque white paper rendered translucent over a small circular region in its centre by the application of paraffin-wax. This sheet S (*Fig.* 6) is mounted in a box which is blackened on the inside and provided with two mirrors MM by means of which the two sides of the sheet may be simultaneously observed. The box is mounted between two sources of light, the sheet of paper being perpendicular to the line joining them, and is moved to and fro along this line until the observer obtains an identical appearance for the two sides of the paper as viewed in the mirrors. The candle-powers of the sources are then in the ratio of the squares of their distances, d_1, d_2, from the paper sheet in the photometer head, *i.e.* if J_1 and J_2 are the candle-powers of the sources in the direction of the photometer, then

$$\frac{J_1}{J_2}=\frac{d_1^2}{d_2^2}.$$

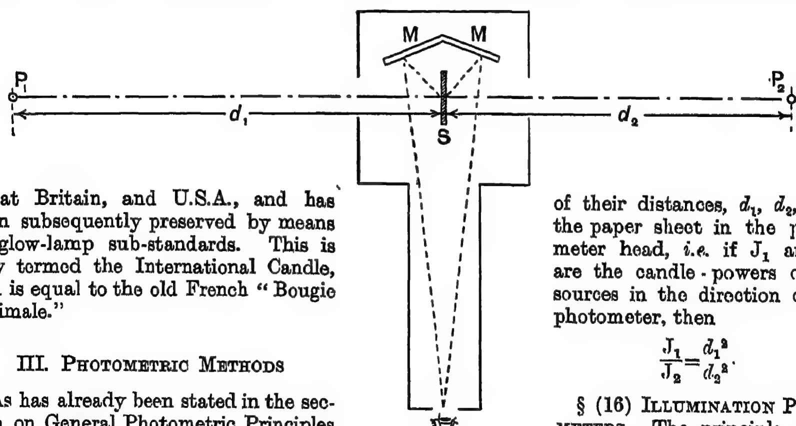


Fig. 6.

§ (16) Illumination Photometers.—The principle of an illumination photometer is somewhat different. This instrument is used for determining the illumination at a given position by placing a matt white surface in that position and measuring its brightness by comparison with that of another surface contained in the instrument. The illumination of this surface is variable at will by the motion

of some part of the instrument, so that, with a scale previously calibrated, the value of the illumination of the outside surface can be at once obtained. *Fig.* 7 shows a typical form of illumination photometer.

P is the photometer case containing L the electric lamp, S_2 the surface whose brightness can be varied by moving L, and E an eyepiece. S_1 is the other surface placed at the point at which it is desired to measure the

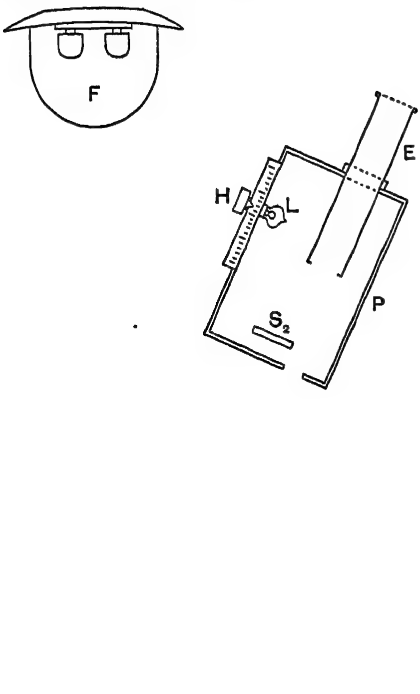


FIG. 7.

illumination produced by lamp F. The surfaces S_1 and S_2 are observed simultaneously through E, and the handle H is moved until S_1 and S_2 appear equally bright. As the movement of the handle regulates the illumination at S_2 each position taken up by it corresponds with some definite illumination. These illuminations are marked on a scale over which moves a pointer attached to the handle, so that when there is equality of brightness the scale reading indicated by the pointer gives at once the value of the illumination at S_1, which was desired.

§ (17) MODERN METHODS OF PHOTOMETRY.—Modern photometry, as far as it is concerned with the comparison of light sourcés, and their measurement in a given direction by comparison with standard sources, depends on the use of a comparison photometer of the most sensitive type available, together with a photometer bench specially designed for the accurate determination of the distances involved.

§ (18) THE LUMMER-BRODHUN PHOTOMETER.—The form of photometer head generally employed for work of the highest accuracy with lights of the same colour is the Lummer-Brodhun contrast type, the first form of which was described in 1889, though it has undergone several improvements since that date. The principle on which the instrument works will be best understood from *Fig.* 8, which shows on the left a plan view of the interior of the photometer head, and on the right an enlarged view of the prism system P. S is a central screen (approximately 4 mm. thick) constructed of white plaster with as matt a surface as can be obtained. The two sides of this screen are respectively illuminated by light from the two lamps to be compared. M_1 and M_2 are total reflection prisms, and by means of these the light from the two sides of the screen is brought to opposite faces of the prism system P.

This system, as may be more clearly seen from *Fig.* 8A, consists of two right-angled prisms placed with their hypotenuses in contact. The hypotenuse of the left-hand prism, however, is sand-blasted with the pattern shown shaded in *Fig.* 8B, and thus the only parts of the two prisms which are in optical contact have the form of the white pattern in that figure. The result is that light passing into prism P_1 is transmitted without change to prism P_2 in the pattern shown white in *Fig.* 8B, while the pattern shown shaded in that figure is the pattern over which total reflection takes place in prism P_2, *i.e.* the pattern over which the light from M_2 is seen by the observer at O. In effect, therefore, the observer at O sees a pattern of this form in which the brightness of the shaded portion is due to light from the right-hand side of S, while the brightness of the white portion is due to light from the left-hand side of S. Clearly, when the two sides of S have the same brightness the pattern will disappear.

Disappearance, however, is not the condition of which the eye is capable of judging most sensitively, and therefore sheets of glass G_1 and G_2 are inserted as shown in *Fig.* 8A, so that the light forming each of the rhomboidal patches is reduced by 8 per cent, and the condition to be arrived at is then equality of contrast between the patch and its background in both halves of the field of view. The observer is provided with a telescope at O, by means of which the pattern of the field is brought into accurate focus for his eye,

as without sharpness of focus it is difficult to obtain accurate settings of the photometer.

The whole of the optical part of the apparatus is mounted rigidly in a brass box which is capable of rotation about a horizontal axis. Further, the screen S can be removed for alignment of lamps on the bench, or for reversal of screen to eliminate differences due to lack of equality in reflection ratio of the two plaster surfaces. The photometer windows are provided with brass shutters, and these should always be kept closed when the instrument is not in use, as otherwise both the plaster screen and the glass surfaces of the prisms become dusty, and the field of view is covered with specks so that the accuracy of the readings is much impaired. With experienced observers, lights of the same colour can be compared with this photometer to an accuracy of 0·1 per cent by taking the mean of a number of observers working on several occasions and using the method of procedure described later in this section.

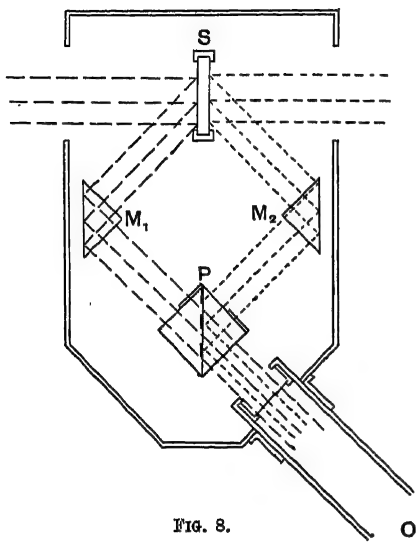


FIG. 8.

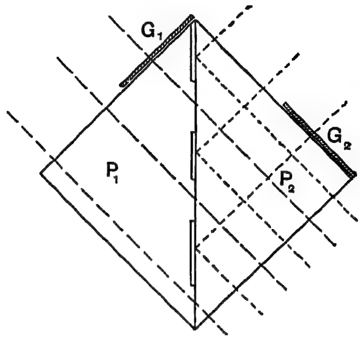


FIG. 8A.

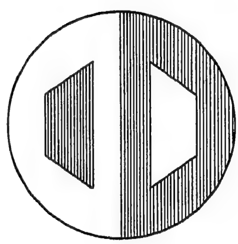


FIG. 8B.

§ (19) THE PHOTOMETER BENCH.—The photometer bench must next be described. The pattern designed and used at the National Physical Laboratory is shown in general view in *Fig.* 9. It consists of two straight steel bars, 3000 to 5000 mm. long, supported rigidly at intervals, the distance between the bars being 300 mm. Carriages, one of which is shown to a larger scale in *Fig.* 10, support the lamps to be compared, and the photometer head. These carriages consist of aluminium base plates B supported on three rollers with V-shaped grooves which run smoothly along the bars of the bench. The centre upright of the carriage has an adjustment for raising and lowering the plate P, and the latter is also capable of rotation about its axis, and carries a scale of degrees, so that it may be brought to any desired position.

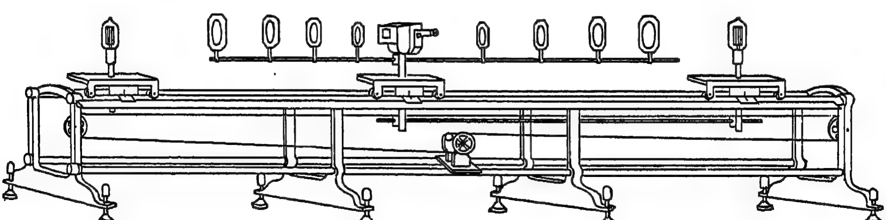


FIG. 9.

Different fittings are attached to the various carriages according to the particular apparatus which they are intended to bear. The speci-

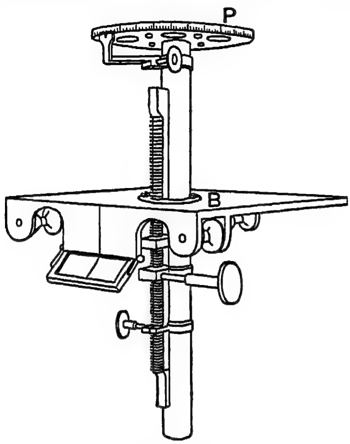


Fig. 10.

men shown in *Fig.* 10 is designed to hold a lamp. Others are adapted for carrying a rotator, apparatus for polar curve measurements, or a photometer head and screening system.

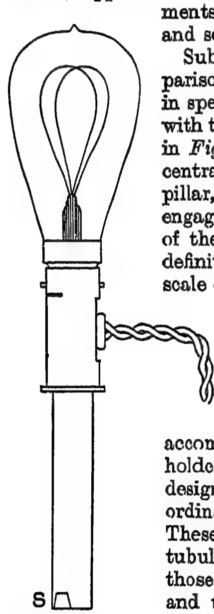


Fig. 11.

Sub-standards and comparison lamps are mounted in specially designed holders, with tubular stems (as shown in *Fig.* 11) which fit into a central hole in the carriage pillar, a slot S at the bottom engaging in a key at the base of the hole, so that when a definite mark on the degree scale of the table is opposite the pointer the lamp is in a definite position with respect to the axis of the bench.

Test lamps are accommodated in special holders which have sockets designed to take lamps with ordinary standard caps. These holders terminate in tubular stems similar to those on the sub-standards, and they are provided with two pairs of leads, both of which are soldered to those parts of the sockets which make the contacts with the lamp cap. One of these pairs of leads is used for supplying current to the lamp, while the second pair is used for voltage measurement. In this way the voltage is measured at points which are as near as possible to the actual lamp contacts, and no allowance has to be made for voltage drop in the supply leads. Each pair of leads is connected at its free end to a special form of ebonite holder, which is designed to facilitate connection with terminals carried on a small ebonite board at the end of the bench (*Fig.* 12).

The carriage bearing the photometer head also carries a steel bar along which are placed,

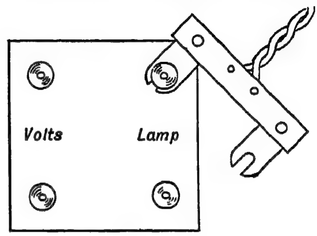


Fig. 12.

at convenient intervals, a number of blackened aluminium screens with varying sizes of apertures. The relative sizes of the screens and their apertures, and the intervals at which they are placed along the bar, are so related that the screen of the photometer head is completely shielded from rays of light proceeding from anywhere but a narrow region surrounding the lamp to be measured (see *Fig.* 13). When the bench is in use black curtains are hung on either side of it throughout its length, and black velvet screens are placed behind the lamps being compared, so that, as far as possible, stray light is completely

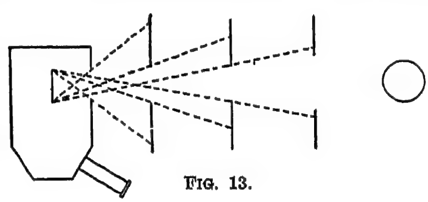


Fig. 13.

prevented from reaching the photometer. One source of stray light which is sometimes found to give trouble is the specular reflection of the light from the lamp by the polished surfaces of the steel rods of which the bench is composed. This may be avoided by ebonising the rods (though even in this case specular reflection still takes place) or, better, by covering the rods at the half-way point with a piece of black velvet. The importance of adequate screening cannot be overestimated in accurate photometric work.

§ (20) Procedure in making a Photometric Measurement.—The photometric procedure adopted at the National Physical Laboratory may be conveniently described by analogy with the double-balance method

of weighing. As in this method the object whose mass is required is first balanced by an unknown mass of sand and is then removed from the scale pan and replaced by weights sufficient to balance the sand again, so in this method of photometry a lamp of unknown candle-power is used on the right-hand side of the photometer bench, and this is termed the "comparison lamp." The carriage holding this lamp is clamped to the carriage holding the photometer head by means of a bar of convenient length, so that the two carriages may be moved as one unit and the illumination on the right-hand side of the photometer head remains constant (see *Fig.* 9).

To adjust this illumination to 10 metre-candles a glow-lamp sub-standard of accurately known candle-power is placed in a carriage at the zero mark on the left-hand end of the photometer bench. The distance of this lamp from the photometer head, which gives an illumination of 10 metre-candles on the screen, being accurately known, the corresponding distance for the comparison lamp (termed its "fixed distance") is found by photometric balance between the two. In actual practice this fixed distance is found as the mean of observations with four or five sub-standards and two or more observers, depending on the accuracy of the work to be undertaken. The distance between the photometer head and the comparison lamp is then fixed, by means of the bar, at the value thus obtained, and the bench is then ready for the measurement of the test lamps.

A test lamp having been placed in the left-hand carriage of the photometer bench, and the axis of its filament (or mean plane of filament when dealing with grid filaments) having been carefully adjusted to be over the zero of the photometer bench, the observer makes a number of settings of the photometer head. An assistant notes these down, and the square of the mean distance in metres multiplied by 10 gives at once the candle-power of the test lamp in the direction of the photometer. It is generally the case that two observers will obtain values for the fixed distance of the comparison lamp differing by several millimetres in 1300 to 1500 (the distance for a lamp giving about 20 candles). It is therefore customary to set the fixed distance at the value found by one observer and to correct the observations of the other observers so as to make these correspond with the fixed distances found by them.

It may be mentioned in passing that the observer at the photometer head never sees his own readings until they are all taken and entered up, so that he has no chance of being unconsciously biassed in one direction or another. For really accurate work, such as the standardisation of sub-standards, where values are certified to an accuracy of a quarter per cent, the procedure above outlined is gone through on three or four separate occasions, as it is found that the relative values obtained by two or more observers will differ slightly from day to day, so that it is desirable to have the mean results of several days' observations.

(i.) *Errors.*—It will be seen that this method of photometry avoids four of the errors to which a simple comparison of test lamp with standard is liable. These errors are: (i.) photometer screen error, (ii.) zero error of photometer head, (iii.) unequal reflections from extraneous objects at the two ends of the bench, and (iv.) observer's personal error. The first of these errors, due to lack of symmetry in the photometer head, as has already been mentioned, can be compensated by reversing the photometer head, but this does not apply to the other errors, which still remain unless a substitution method is employed.

It will be clear that the bench, in addition to its millimetre scale, may bear a "square" scale so graduated that the position of the photometer head, when the illumination is equal to 10 lux, gives the candle-power of the test lamp directly without calculation; for with a reading of n millimetres the candle-power is $(n/1000)^2 \times 10$ candles. It will be noticed that no allowance has been made above for the thickness of the plaster screen of the photometer head. If the distances of the two lamps from the photometer are approximately equal, or even in the ratio of 2 to 1, with distances of 1300 mm. or over, the error introduced by this neglect does not exceed 0·15 per cent. For work at short and unequal distances, however, the semi-thickness should be subtracted from the distance of the test lamp, assuming that the distance of the sub-standard has been given as that necessary to produce an illumination of 10 metre-candles on the actual surface of the photometer screen.

(ii.) *Alternative Method.* — It sometimes happens that the method of fixed distance described above is not practicable, either on account of the high candle-power of the test lamp and insufficient length of the bench, or when measuring a source in a number of positions in which the candle-power varies over a wide range. In this case it is necessary to fall back on actual candle-power measurements of the comparison lamp, and the use of the inverse square law, both lamps being held stationary, while the photometer head alone is moved. In this case, if d is the distance between the two lamps, J the candle-power of the comparison lamp, and x the distance of the photometer head from the test lamp, the candle-power of the test lamp is given by the formula $Jx^2/(d-x)^2$. This method, of course, involves

much more calculation than the fixed distance method.

For the measurement of very high candle-power sources at the National Physical Laboratory, a 3-metre photometer bench is mounted on a table fitted with rollers, which move along a rail track 30 metres in length.

§ (21) BEST ILLUMINATION.—An important factor which has to receive consideration in accurate photometry is the degree of illumination desirable on the photometer screen. It seems to be generally agreed that an illumination of between 5 and 20 metre-candles is that at which the eye is capable of giving the best results with the Lummer-Brodhun photometer head. Outside these limits the accuracy of judgment of equality begins to diminish.

(i.) *Absorbing Screens.*—In the case of very high candle-power sources of light it is sometimes inconvenient or impossible to place these sufficiently far from the photometer to give the desired degree of illumination, and various methods have been proposed for reducing the intensity in a determinable ratio. One such method, the use of a neutral-tinted glass plate (or double-wedge), placed in the path of the light, has the disadvantage that truly neutral glass is unobtainable, so that in practice it is necessary to determine the transmission ratio of a given specimen of glass absorber by means of light of exactly the same colour as that with which it is intended to be used. An alternative is that proposed by Ives and Luckeish,[1] and elaborated by Krüss.[2] This consists of two glass plates ruled with fine opaque black lines, exactly equal in breadth to the spaces between them. These two plates slide one behind the other, and so a variation in transmission of 50 per cent to zero can be obtained.

(ii.) *The Sector Disc.*—The most generally used apparatus for the purpose, however, is the sector disc in one of its many forms. The pattern devised and used by Abney is shown in *Fig.* 14, and possesses the advantage that the angle of the sector openings can be varied while the disc is in motion. The disc is placed so that its upper portion alternately intercepts and transmits the beam of light which it is proposed to reduce. The shaft carries near one end a grooved pulley driven at any desired speed by an electric motor. At the other end is a disc A, of which three equal sectors have been removed, except by the shaft and the rim. A second, exactly similar, disc is placed behind this one, and is rigidly attached to a flange fixed to a sleeve which slides on the shaft, and has a pin engaging in a spiral groove cut in the shaft. Thus, the longitudinal position of the sleeve along the axis of the shaft controls the relative positions of the two discs, and so the width of the sector openings is capable

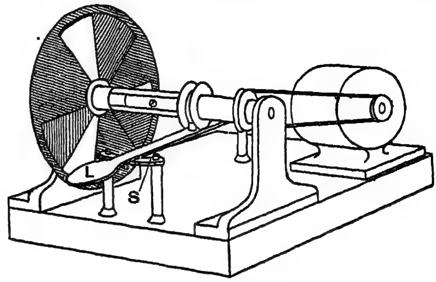


FIG. 14.

of control by means of a grooved wheel attached to the sleeve and acted upon by a pin in a lever L, which moves over a divided scale S. It has been shown by E. P. Hyde[3] that the transmission ratio of a disc of this kind, if due precaution be taken to avoid stray light, is accurately the same as the ratio of the total angle of opening to 360°. Of course, the smaller the opening the more the accuracy of the transmission ratio depends on the accuracy with which the sectors are cut, and for this reason it is not generally advisable to use openings smaller than 10° with a disc of ordinary construction. The speed of rotation has to be adjusted until all flicker of the field disappears, and therefore needs to be higher the smaller the transmission ratio.

§ (22) ERROR DUE TO SIZE OF SOURCE.—An important consideration in the photometry of sources of light of large dimensions, or where the candle-power is so small that the distance from the photometer has to be made comparable with the dimensions of the source, is the limit at which the inverse square law may be taken to apply with the necessary accuracy. This law is, of course, only strictly applicable to a point source of light, and in the practical case of a source of finite dimensions, the illumination of the photometer screen is the sum of the partial illuminations due to all the elementary portions of which

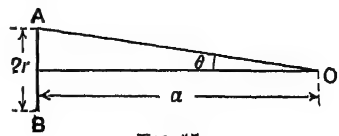


FIG. 15.

the source is composed, the inverse square and cosine laws being applied to each such elementary portion separately. Thus, in the case of a circular disc AB (*Fig.* 15), the illumina-

[1] *Phys. Rev.*, 1911, xxxii. 522.
[2] *Zeits. Instrumentenk.*, 1917, xxxvii. 109.
[3] *Bureau of Standards Bull.*, 1906, ii. 1.

tion of an elementary surface at a point O along the axis of the disc, due to an element of the disc of area a situated at its centre, is aI/a^2, where I is the normal brightness of the disc. If the disc is a perfect diffuser, the amount of flux it emits in any direction is proportional to the cosine of the angle which that direction makes with the normal to the surface. The illumination produced by a similar element at A, therefore, is only $aI\cos^2\theta/AO^2$, for the amount of flux emitted by the disc per unit area in the direction of AO is $I\cos\theta$, and since this meets the surface at O at an angle θ with the normal to that surface, the illumination is again subject to the factor $\cos\theta$. Hence the illumination at O due to A is $aI\cos^4\theta/a^2$. By integrating this expression over the whole disc it is found that the illumination at O due to the whole disc is $r^2E/(a^2+r^2)$, where r is the radius of the disc and E is the illumination calculated on the assumption that the size of the disc is negligible in comparison with its distance from O. Similarly it may be shown that for a single straight filament of length $2l$ the illumination at an elementary surface distant a from its centre is $E/2[a/l\tan^{-1}l/a+a^2/(a^2+l^2)]$, where E has the same meaning as before. Clearly, in these two particular examples, if the error is not to exceed 0·2 per cent, then in the first case r/a must not exceed 4·5 per cent, and, in the second case, l/a must not exceed 4 per cent. In *Fig.* 16 are given graphs of the percentage errors introduced by assuming discs or lines of various dimensions to behave as absolute point sources. These graphs give therefore the dimensions of the largest sources for which the inverse square law may be assumed to hold to any desired degree of accuracy.

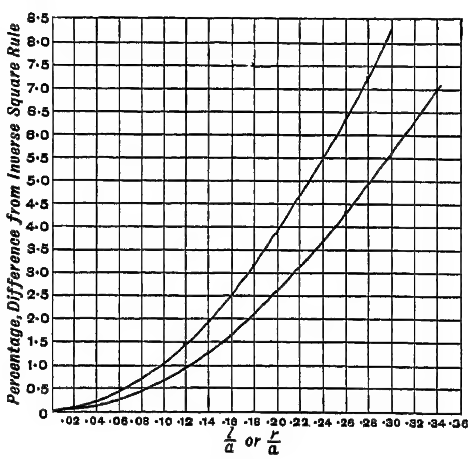


FIG. 16.

§ (23) VARIATION OF CANDLE-POWER WITH VOLTAGE.—As has been already stated above, the candle-power of electric glow lamps varies at a much more rapid rate than the voltage applied to the filament. Actually it has been found that for tungsten filament vacuum lamps, a voltage change of 1 per cent causes a 3·7 per cent change of candle-power, while for carbon filament lamps this change is as much as 5 per cent. For tungsten filament gas-filled lamps the figure is generally not much different from that for tungsten filament vacuum lamps. For tungsten the change produced by a given current variation is approximately half that produced by the same percentage change of voltage.

(i.) *Voltage Regulation.*—From this it will be seen that to attain an accuracy of one-tenth per cent in candle-power measurements it is necessary to ensure that the electrical measurements and regulation shall be accurate to at least 0·02 per cent. Either voltage or current regulation may be employed; the latter has the advantage that it is not necessary to ensure that the electrical measurements are made at the terminals of the lamp, but the former method has the advantage of greater sensitivity, and is the method generally employed in photometric laboratories.

(ii.) *Voltage and Current Measurement.*—For work where an accuracy of 1 per cent in candle-power is the best aimed at, indicating instruments of a large scale precision type are good enough if constantly checked against a standard cell and accurate potentiometer. For more accurate work, however, a potentiometer method of voltage measurement must be employed, and it is essential that the voltage of the supply shall be absolutely free from momentary fluctuations. In the most accurate work a storage battery, of reasonably high capacity for the loads to be taken, is essential. The leads from this supply are brought through adjusting resistances to the current terminals on the photometer bench. From the voltage terminals, leads are carried to the terminals of a potentiometer, which is repeatedly checked during the course of a day against a standard Weston cell. If measurements of the current passing through the lamp are also desired, it is necessary to introduce into the main circuit of the lamp an accurately measured resistance capable of carrying the current without sufficient change of temperature to affect the value of the resistance. The voltage across the ends of this standard resistance can then be measured by means of the potentiometer and the value of current deduced.

Frequently, when using two electric lamps on the bench at the same time, it is convenient to be able to have a constant indication of the voltage on each lamp, and in this case an electrostatic voltmeter may be usefully employed on the comparison lamp circuit. This lamp has normally to be run for a considerable length of time at a constant voltage, and therefore a voltmeter with a sufficiently enlarged scale (that used at the National Physical Laboratory has a scale of 12 feet radius on which one volt is represented by a length of 2½ inches)[1] may be used for maintaining a watch on its voltage. The indication of this voltmeter has to be checked at intervals throughout the day on account of the slow upward creep due to the lag of the suspension. With this arrangement, the potentiometer is

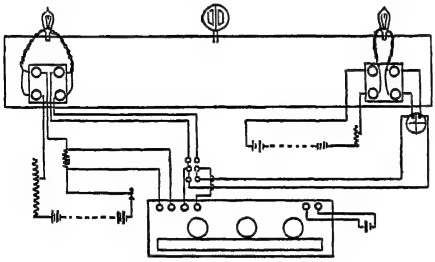


FIG. 17.

free to give a constant indication of the correctness of the voltage on the test lamp or sub-standard. A sketch diagram of the electrical connections is given in *Fig.* 17.

§ (24) OTHER TYPES OF PHOTOMETER.—The above is a general description of the methods usually followed when using a comparison photometer of any ordinary type and for the purpose of illustration; the Bunsen grease spot and the Lummer-Brodhun contrast photometers have been described. The method, however, is perfectly independent of the particular form of photometer head employed, and some of the other types which have been devised must now be described. The total number of these is very large, and it is impossible to do more than give a brief outline of a few of the best known in the space here available. For the others, a text-book on photometry, such as that of Liebenthal, *Praktische Photometrie*, or Trotter, *Illumination, Its Distribution and Measurement*, should be consulted. All of them depend, of course, on the comparison of brightness, but the arrangement of the surfaces to be compared and the form of the line of separation differ in the different instruments. Further, while the majority depend on the law of inverse squares for the variation of the illuminations, some use other means, such as polarisation, for this purpose, so that in these latter instruments the sources and the photometer are not altered in relative position while the measurements are in progress.

Of the first class, depending on the inverse square law, the Rumford, the Harcourt Gas Referees, the Ritchie, Conroy, Thomson-Starling, Joly, and Trotter types will be described, while the Martens will be taken as a model of the polarisation type.

§ (25) THE RUMFORD PHOTOMETER.—The earliest form of photometer capable of accurate work was designed by Count Rumford in 1793, and consisted of two vertical cylindrical rods R_1 R_2 (*Fig.* 18), contained in a blackened box at the meeting-point of two tables. The tables carried the two sources to be compared, and these cast shadows of the cylinders on a white surface at the back of the box. The relative positions of the cylinders were so adjusted that the two central shadows were just in contact at the centre of the white surface, while the two outer shadows were cast on a blackened surface and so lost. An opening in the centre of the front of the box enabled an observer to compare the intensities

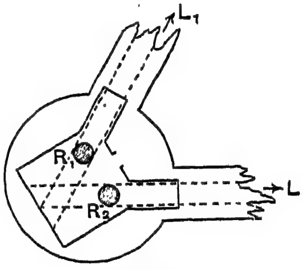


FIG. 18.

of the two shadows, and to obtain equality between them by moving one of the sources.

§ (26) THE HARCOURT PHOTOMETER.—The Harcourt photometer used by the Metropolitan Gas Referees consisted, in principle, of a box, having on one side a circular opening covered with translucent paper and on the other a diaphragm with a rectangular opening 25 mm. high and 7 mm. broad. The two sources of light to be compared were placed on the diaphragm side of the box, each slightly to one side of the line joining the diaphragm to the screen (see diagram *Fig.* 19). In this way each source caused the diaphragm to throw a bright rectangular patch on the screens, and by suitably altering the positions of the lights these two patches could be brought into juxtaposition, so that their brightness could be readily compared. When equality of brightness was obtained, the candle-powers of

[1] "Photometry at the National Physical Laboratory," *Illum. Eng.*, London, 1908, i. 845.

the two sources were in the ratio of the squares of their distances from the screen S. It was, of course, essential that the normal to the screen

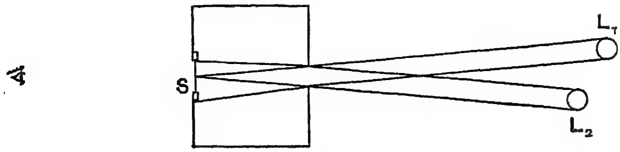


FIG. 19.

should exactly bisect the angle between the lines joining the centre of the screen to the two light sources, and in the actual instrument a mirror attachment was provided for the purpose of making this adjustment.

§ (27) THE RITCHIE WEDGE.—The photometer designed by Ritchie in 1826 consists of two inclined mirrors MM (see *Fig.* 20). These reflect the light from sources L_1, L_2, to a piece of translucent paper P. If the angle between the mirrors be rather more than 90°, the bright patches produced on the translucent paper may be brought into accurate contact, and then balance is obtained in the same way as with the Bunsen photometer by altering the distance of one or both lamps until equality of brightness between the patches is obtained. In a modification of this photometer matt white surfaces are used in place of the mirrors, and the translucent paper is removed, so that the surfaces of what has been termed the "Ritchie wedge" are directly viewed by the observer.

A very important factor in all photometers is the accurate placing in exact juxtaposition without interval or overlapping (as far as the eye is concerned) of the surfaces to be compared. Even a fine line of separation, either darker or brighter than the surfaces compared, impairs the accuracy of equality judgment, though this is of less importance in a contrast type of field. With the Ritchie wedge, therefore, the front edge must be exceedingly fine, and further, on account of the cosine law of illumination, it is most important that the light shall impinge

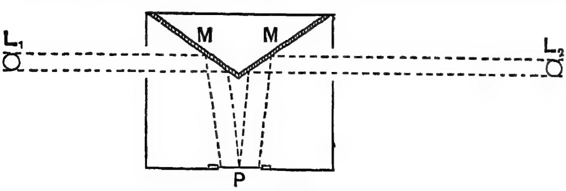


FIG. 20.

on both surfaces at exactly the same angle. Trotter has shown,[1] however, that for wedges made of dulled Bristol board the error introduced by shift of angle is not nearly as great as would be expected from the simple cosine law, owing to the special behaviour of this material when the light is incident at an angle of 30° to 35°. An angle of 60° to 70° is therefore best if the Ritchie wedge is made of this material.

Two modified forms of the Ritchie wedge photometer have been used. The first of these was introduced in 1883 by Conroy and the

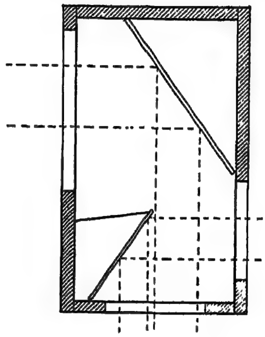


FIG. 21.

second by Starling and S. P. Thompson in 1893. They are shown in *Figs.* 21 and 22 respectively, and from these diagrams the

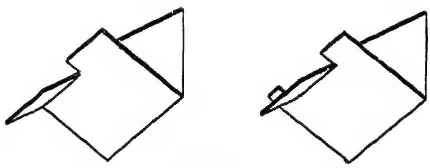


FIG. 22.

principles of the designs are self-evident. To ensure a sharp line of junction in the Thompson-Starling pattern, the edges of the cards facing the observer should be sharply bevelled.

§ (28) THE TROTTER PHOTOMETER.—Trotter's "Perforated Disc" photometer is shown in plan in *Fig.* 23 The two inclined cards A and B are of Bristol board with the glaze removed by passing over it a damp rag. A has a star-shaped hole cut in its centre, the length and height being so proportioned, and the edges bevelled so that the observer at O sees a perfectly sharp-edged and symmetrically

[1] *Illumination, Its Distribution and Measurement,* p. 95.

shaped cavity behind which appears the surface of B. A and B being illuminated by the two sources, it is possible to obtain an accurate

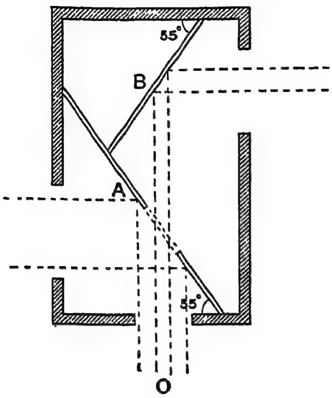


FIG. 23.

balance of brightness between them when the edge of the star becomes almost invisible. A similar design, independently proposed by Weber, is termed the "Roof" photometer, from the appearance of the cards when seen from above.

§ (29) THE JOLY PHOTOMETER.—The photometer designed by Joly in 1888 consists of two blocks of paraffin wax, or translucent glass, placed side by side. Light from the two sources falls upon the side of the blocks which appear to be suffused with light, and the brightness match is obtained between the two sides B_1, B_2 (*Fig.* 24). For an illumination of 10 lux with paraffin wax as the material, blocks about 5 mm. thick give a convenient brightness. If thicker blocks are used the illumination must be increased. When nearly balanced the less brightly illuminated block appears to have a grey band next to the dividing line, and as the point of balance is passed this band appears to shift from one side of the line to the other. The point at which this shift appears to take place gives a very sensitive criterion of the exact position of balance.

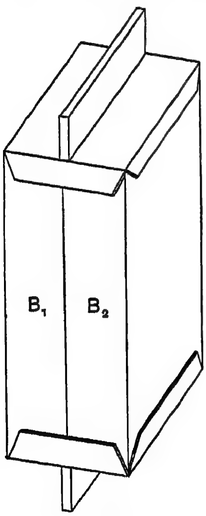


FIG. 24.

§ (30) THE MARTENS POLARISATION PHOTOMETER.—The chief example of a photometer employing polarisation as the means of varying the brightness of the comparison surfaces is Martens' Polarisation Photometer, shown in diagrammatic section in *Fig.* 25. In this instrument the light from the test lamp illuminates a plaster screen F. The light from this is reflected by two right-angled prisms P, Q through the Wollaston prism W, and the bi-prism B to the analyser N. Light from a small comparison glow lamp G illuminates an opal glass plate *b*, and this also is seen after transmission through W, B, and N. W and B are so arranged that the light from the opal plate *b* which reaches the eyepiece is polarised at right angles to that by which F is seen. Hence, equality being obtained by rotating the Nicol N, the ratio between the candle-powers of any two lamps placed successively at equal distances from F is the same as the ratio of the square of the tangents of the angles of rotation of N from the zero position. The tube T is capable of rotation about the axis passing through the centres of the prisms P, Q.

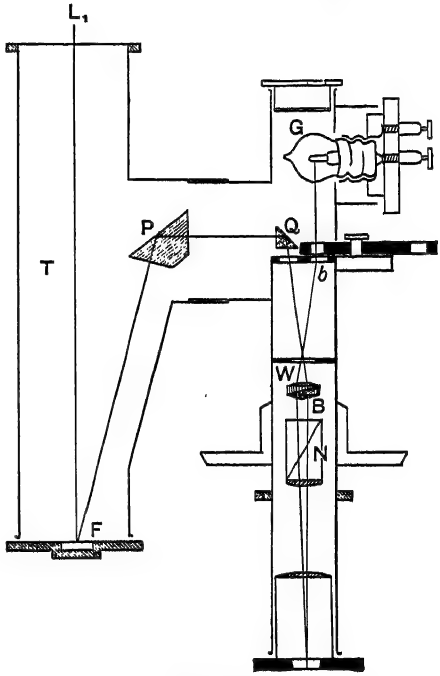


FIG. 25.

§ (31) THE ACCURACY OF MEASUREMENT.—The relative merits of the different patterns of photometer are very difficult to decide. The accuracy obtainable in photometric work depends very largely upon the individual and

each observer will obtain the best results with the particular instrument to which he is accustomed. Even the same observer differs in accuracy from day to day, and with the extent to which his eye has been fatigued by previous work. The time taken in making a reading differs with different observers, but if too long a time is taken, the precision of judgment tends to diminish after the first twenty seconds or so. The part of the apparatus which has to be moved to obtain a balance should be light enough to require very little manual effort on the part of the observer. It is often found convenient to approach the position of balance by overshooting on each side a number of times in succession, the amplitude of the overshoot being diminished each time. If the eye is allowed to see anything brighter than the field of view in the photometer, its power of accurate balance is destroyed for a time depending on the brightness of the object seen and the time for which the eye has been exposed to it.

§ (32) Physical Photometers. — All the photometers hitherto described have depended on the comparison of brightness by the human eye. It is self-evident that, in the case of all measurements of light, the eye must be the final judge of equality, but since individual eyes differ slightly from one another in their judgment it is inevitable that the results obtained by what may be called physiological photometry cannot be independent of the peculiar characteristics of the observer, and many proposals have been made to place photometry on a semi-physical basis—*i.e.* to design some instrument which will respond to light in the same way as the "normal eye" or the average of a very great number of individual eyes none of which possesses any marked abnormality as regards light perception.

The instruments proposed for the purpose of physical photometry may be grouped in three classes according as they depend on (i.) change of the electrical properties of a material when illuminated, (ii.) the electron emission of a metal surface under the action of light, or (iii.) the action of light radiation on a thermopile or bolometer.

§ (33) The Selenium Cell.—Of the first class, the most commonly used substance is selenium. It has long been known that the electrical resistance of selenium falls when illuminated, and many devices for the automatic lighting of lamps depend on the rise of resistance of a selenium cell when the daylight illumination falls.

One form of selenium cell may be made by taking a small sheet of ground glass and spreading this with a very thin layer of purified amorphous selenium by means of a hot glass rod. If now four strands of fine bare copper wire are wound round the plate so that the whole of the selenium surface is covered, and two alternate strands are then removed, the other two strands are left separated for the whole of their length by a space equal to the diameter of a wire. To make the cell sensitive to light it is heated in an oven to a temperature of 180° C. for about five minutes, when the transformation from amorphous to metallic selenium should be complete. The resistance of such a cell, formed on a plate $1 \times 3 \times 0{\cdot}1$ cm., has been found by Pfund [1] to be of the order of 2×10^7 ohms, and the sensitivity such that the resistance is decreased to 10 per cent of this value by an illumination of 150 to 200 metre-candles. The cells should be protected from moisture by being placed in a vacuum tube or waxed to a sheet of glass or mica.

A selenium cell is very selective, having a maximum of sensitivity at a wave-length of about 700 $\mu\mu$. The most sensitive cell made by Fournier d'Albe [2] was capable of detecting an illumination of the order of 10^{-5} metre-candles. He found that the change of resistance of a cell was, for a given period of recovery, approximately proportional to the square root of the illumination.

The chief disadvantage attending the use of a selenium cell is its slowness in recovering its resistance after exposure to illumination, the period taken for recovery increasing with the intensity of the illumination to which it has been exposed.

§ (34) The Photo-electric Cell. — The second class of physical photometer, generally called the "photo-electric cell," depends upon the fact, discovered by Hallwachs in 1888, that a body carrying a negative charge of electricity loses that charge when ultra-violet light falls upon it. Elster and Geitel in 1889 [3] showed that sodium and potassium exhibit this same effect when exposed to light in the visible spectrum, and they and other workers have devoted much careful experimental work to the study of this effect and the possibility of its application to physical photometry.

A convenient form of the apparatus for fairly high illumination intensities is shown in *Fig.* 26.[4] The photo-electric cell itself consists of a glass globe Z containing a mixture of helium and argon at low pressure. Into the centre of this cell projects a platinum wire which acts as the anode of the cell and is connected to the terminal K_2 of the photometer. The cathode consists of a surface of potassium which makes contact with a silver plate Ag and is connected by means

[1] *Phys. Rev.*, 1909, xxviii. 324.
[2] *Roy. Soc. Proc.*, 1913–14, lxxxix. 75.
[3] *Annal. d. Physik.*, 1889, xxxviii. 40, 497.
[4] Elster and Geitel, *Phys. Zeit.*, 1912, xiii. 740.

of a platinum wire sealed into the glass with the terminal K_1 of the photometer. A guard-ring S_1 connected to earth prevents

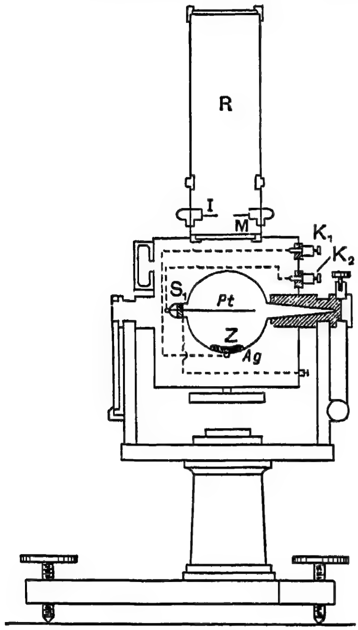


FIG. 26.

leakage over the surface of the glass between the two electrodes. The light enters by the tube R, which is closed by a cap when the photometer is not in use, so that the cell is completely shielded from light except when in operation. I is an iris diaphragm, and M a plate of matt uviol glass. The box containing the cell is blackened inside and is capable of rotation about the axis defined by the platinum wire Pt. Thus the whole instrument is mounted like a theodolite, and the tube R can be oriented in any desired direction. Using a potential difference of 40 to 60 volts between the electrodes on the cell, the current produced by sunlight illumination is of the order of 10^{-6} amperes. The current due to an illumination of 1 metre-candle is of the order of 10^{-11} amps., and it has been found that if proper precautions are taken to eliminate possible disturbing factors very accurate proportionality exists between current and illumination over great ranges of intensity of the latter (0·07 to 6000 metre-candles).[1] It has been calculated by H. S. Allen that an alkali cell of the most sensitive type, used in connection with a tilted electroscope capable of measuring currents of 10^{-15} amperes, should be capable of detecting the light from a candle at a distance of 2·7 miles.[2]

(i.) *Variation of Sensitivity with Wave-length.*—One of the chief difficulties in the use of photo-electric cells as physical photometers is the difference between the behaviour of such a cell and the human eye as regards response to light of different wave-lengths. Normally, the sensitivity of the cell should be greater the shorter the wave-length of the light, but although the alkali metals are selective in the visible spectrum, their characteristic curve is far from even approximating to that of the eye, so that the photo-electric cell in its simple form can only be used for the comparison of lights having exactly the same spectral distribution. In order to make the measurements obtained with them of absolute photometric value, it would be necessary to introduce some colour filter to reduce their curve of sensitivity at different wave-lengths to approximate equality with the sensitivity curve of the average eye.

Various metals and alloys have been tried in different gases and mixtures of gases. Both sodium and potassium have been used separately and alloyed, and the hydrogen at first employed has been replaced by a mixture of helium and argon. Elster and Geitel have prepared very sensitive cells by first heating the alkali metal in hydrogen to a temperature of about 350° C. The clear colourless crystals of hydride thus obtained are then bombarded with cathode rays, large quantities of hydrogen are evolved, and the hydride becomes brightly coloured. These coloured substances, which Ester and Geitel consider to be the metal in a colloidal state dissolved in the solid hydride, are from three to four times as sensitive as the pure metal surface prepared by distillation. The hydride crystals have to be removed from contact with hydrogen, and this is replaced by argon or helium after the colouring has been produced.

(ii.) *Photo-electric Fatigue.*—The second disadvantage attending the use of photo-electric cells is the possibility of loss of sensitiveness by what is known as "photo-electric fatigue." This fatigue may take place when the cell is standing idle and is fairly rapid in the presence of a gas. Much work has been done on both the theory and the practical effects of this fatigue when the pressure of the gas is reduced to that generally used in photo-electric cells (1 to 3 mm. of mercury), but no conclusive evidence of complete absence of fatigue in any particular form of cell has been produced so far.

§ (35) RADIOMETERS.—The third class of physical photometers, depending on radiometric measurements by some instrument such

[1] Richtmyer, *Phys. Rev.*, 1909, xxix. 71, and 1910, xxx. 385.

[2] H. S. Allen, *Photo-electricity*, p. 74.

as the thermopile, has been developed by Ives and Kingsbury in America.[1]

The Thermopile.—It is clear that the thermopile, since it measures total radiation, cannot be used alone to evaluate luminous flux, but that some auxiliary device is necessary to reduce the radiation in each wave-length in such a proportion that the relative effects of the different wave-lengths on the combined apparatus may be the same as their relative effects on the human eye. An approximation to this condition may be arrived at by means of a coloured solution which transmits such a proportion of the whole incident radiation in each wave-length that the effect on the thermopile is proportional throughout the spectrum to the effect of light of that particular wave-length on the human eye. The solution arrived at empirically by Ives and Kingsbury (*loc. cit.*) is as follows:

Cupric chloride . . .	60·0 gm.
Cobalt ammonium sulphate .	14·5 gm.
Potassium bichromate . .	1·9 gm.
Nitric acid (1·05 sp. gr.) . .	18·0 c.c.
Water to	1 litre.

The form of apparatus used by them is shown in *Fig.* 27. T is the thermopile, consisting of

from the Hefner lamp to an ordinary tungsten filament vacuum lamp, the agreements between the visually and physically determined values of transmission ratio of yellow and blue test solutions were identical to one-half of 1 per cent.

As an alternative to the solution method of reducing total radiation to agree with the sensitivity curve of the eye, Ives has proposed[2] the analysis of the light from the source into a spectrum band, by means of a prism. This spectrally distributed light is then transmitted through a sector disc having apertures cut in it of such a shape that the relative amounts of each wave-length transmitted by the disc are proportional to the sensitivity of the eye at that wave-length. The apparatus is shown in plan and elevation in *Fig.* 28, where L is the light source, S the slit, and P the prism producing a spectrum S′. D is the sector disc, and it will be seen from the end view, with one aperture passing over the spectrum, that the different wave-lengths can be

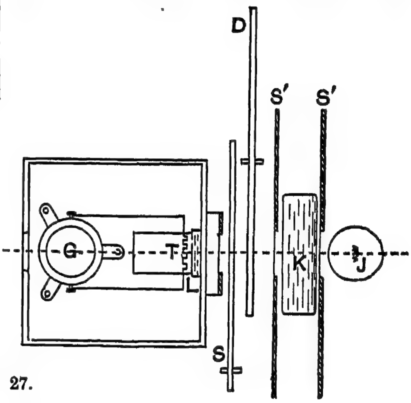
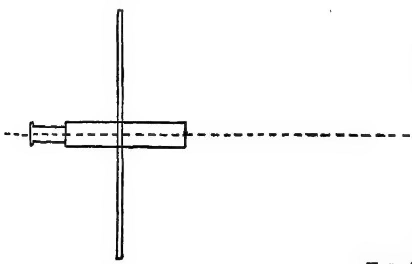



FIG. 27.

18 bismuth-silver junctions arranged linearly in series, and having a total resistance of 26 ohms. G is a d'Arsonval galvanometer with an internal resistance of 12·5 ohms and critical damping resistance of 32·5 ohms. Its sensitivity is 33 mm. per microvolt. L is the luminosity curve solution (above described) in a glass tank one centimetre thick. K is a protective water tank at least four cm. thick, to cut off infra-red radiation. S is a shutter, S′, S′ tin screens, and D a sector disc. J is the source of light, a lamp of 100 candle-power placed 25 cm. from the thermopile. It was found that with this apparatus, the lamp running at the colour of a 4 watt per candle carbon filament lamp and giving a candle-power of 45 candles, the deflection on the galvanometer was 7 cm. It was found that over the range of colour represented by the commoner illuminants and standards, *i.e.*

weighted in such a manner that the sensitivity curve of the eye is reproduced on the thermopile T when the transmitted spectrum is recombined by the lenses C, C.

The chief difficulty met with in the use of total radiation is that, owing to the very small proportion of the total radiation from the source which is in the visible spectrum (generally less than 1 per cent), the effect of scattered heat radiation is very serious, and special precautions have to be taken to ensure that this error is avoided. For details of the methods by which it has been sought to do this, the original papers above referred to should be consulted.

§ (36) SPECIAL PHOTOMETERS.—Descriptions of photometric apparatus specially designed for particular branches of photometric work (*e.g.* the flicker photometer for heterochromatic photometry, the Ulbricht

[1] *Phys. Rev.*, 1915, vi. 319.

[2] *Phys. Rev.*, 1915, vii. 334.

globe for the measurement of average candle-power, and the various forms of illumination photometers) will be found under the sections dealing with those particular branches of photometry.

IV. Light Distribution

The section on Photometric Methods dealt only with methods of measuring the candle-power of a source in one single direction, but in the case of all sources met with in practice, the candle-power varies in different directions, and it is therefore impossible to compare the performance of two sources by a single candle-power measurement unless the direction in which that measurement is made has been carefully defined. It is partly for this reason that sub-standard electric glow lamps are constructed with the filament in one plane. The candle-power in the line perpendicular to this plane is the one taken for purposes of measurement, and this arrangement has the great advantage that a slight departure from the true position has practically no effect on the candle-power in this region.

§ (37) Polar Diagrams.—It is, however, obvious that in the case, for example, of a vacuum glow lamp with a filament of the ordinary squirrel-cage type, the candle-power in the direction of the axis of the lamp is very much less than that measured in a direction perpendicular to this. It is actually found that the candle-power is by no means the same in all directions perpendicular to the axis, and *Figs.* 29 and 29A illustrate these statements. These diagrams are what are generally termed "polar diagrams of light distribution." In these curves the length of the radius vector at any angle gives the candle-power at that angle, and the curve of *Fig.* 29 shows the variation in candle-power at the different angles in a plane passing through the axis of the lamp, while *Fig.* 29A gives similar information for a plane perpendicular to this axis.

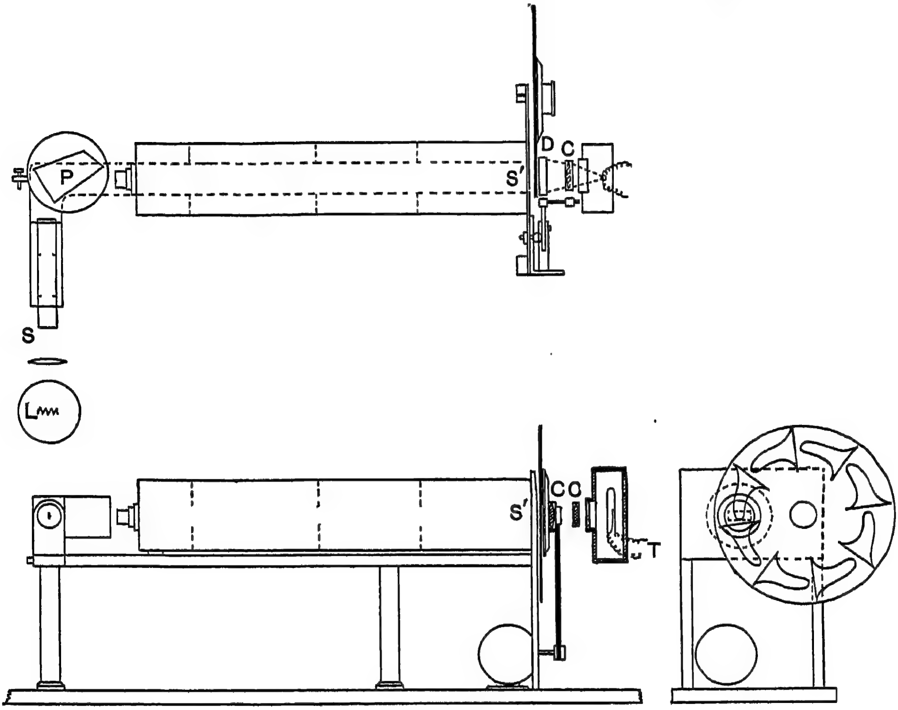


Fig. 28.

§ (38) Mean Horizontal Candle-power.—The different methods by which these polar curves are obtained will be dealt with in this section, but first it is necessary to describe a method for finding by a single measurement the average value of the candle-power in all directions in a plane perpendicular to the lamp axis. This average, called the mean horizontal candle-power (m.h.c.p.) of the lamp, is a figure very frequently used for the rating of vacuum lamps of the ordinary type.

The method consists of rotating the lamp

about its axis (placed so as to be vertical) sufficiently fast to obtain on the photometer head an average illumination. A speed of 120 revolutions per minute is generally

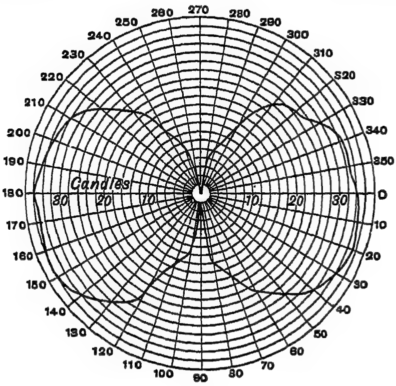


Fig. 29.

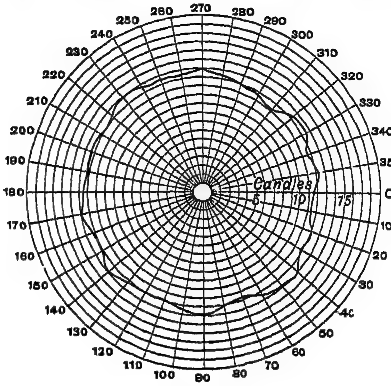


Fig. 29A.

sufficient. The candle-power corresponding to this illumination is then called the mean horizontal candle-power of the lamp. The chief difficulty of this method is to ensure a steady contact for current supply to the rotating lamp when this is being driven sufficiently fast to eliminate flicker in the photometer. This difficulty has been overcome in the form of rotator, shown in *Fig.* 30,

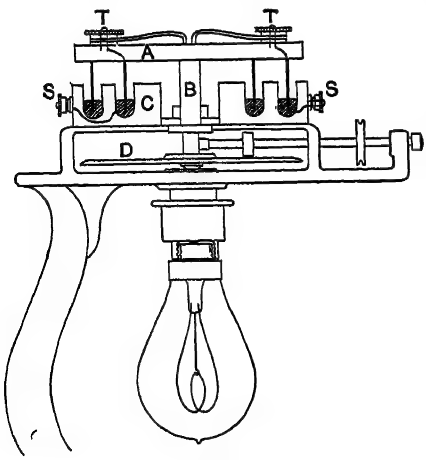


Fig. 30.

which was designed by C. C. Paterson for this purpose.

(i.) *The Lamp Rotator.*—The rotator consists essentially of two parts, one fixed rigidly to the stand, the other capable of rotation and carrying the lamp. The arrangement is best seen from the sectional diagram of *Fig.* 30. A is an ebonite disc bearing two concentric copper rings, which are respectively connected to the terminals T, T. From these terminals flexible leads are conveyed inside the vertical tube B to the lamp holder. Rigidly attached to this tube is a solid brass disc D which is friction-driven by a small wheel driven through a shaft and pulley system from a small electric motor. C is a second circular ebonite block bearing two annular grooves concentric with the copper rings of A and half filled with mercury so that these rings can rotate freely with about ¼-inch of their lower edges dipping into the mercury. The mercury in these two grooves is connected to four terminals S, S to which are attached the current supply leads and voltage measuring leads from the bench. With this apparatus, if the mercury troughs and copper rings are kept clean, there is generally no difficulty in maintaining the current through the lamp perfectly steady.

(ii.) *Step-by-step Method for Gas-filled Lamps.*—The above method of obtaining the mean horizontal candle-power can only be applied to vacuum lamps, as a gas-filled lamp may change candle-power by several per cent if it be rotated.[1] In the case of these lamps, therefore, the method of measurement at different positions round the lamp must be employed. For finding the polar curve in the horizontal plane, *i.e.* the plane perpendicular to the axis of the lamp, supposed upright or pendent, it is only necessary to rotate the lamp and its holder in the photometer carriage and take candle-power measurements every 10°. The correct positions are easily found by the aid of the degree scale on the table fitted to the carriage (see *Fig.* 10). For the determination of the polar curve in a vertical plane the lamp may be mounted

[1] Middlekauff and Skogland, *Bureau of Standards Bull.*, 1915–16, xii. 595.

in the fitting shown in *Fig.* 31. This consists of two semi-rectangular frames of iron, jointed at the open ends, so that one frame can be clamped at any desired angle with the other. A graduated dial is fixed to the lower frame (which is made to fit into a photometer carriage) so that the angle of tilt of the upper frame, bearing the lamp, can be ascertained. By this means a curve showing the distribution of candle-power of a lamp, in a plane passing through its axis, can be obtained. A curve thus obtained for an ordinary "squirrel cage" tungsten-filament vacuum lamp is shown in *Fig.* 29A.

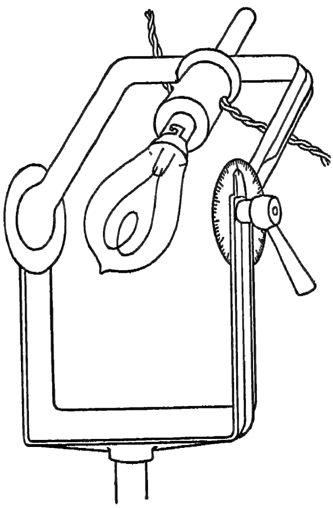


FIG. 31.

§ (39) THE MIRROR APPARATUS FOR POLAR CURVE DETERMINATION.—It is not possible, however, to tilt all light sources. Gas flames or mantles clearly cannot be tilted, and some gas-filled electric lamps change candle-power appreciably according to the position in which they are burning.[1] For such sources, therefore, it is necessary to have recourse to some arrangement of mirrors. A simple form of apparatus for this purpose is shown in *Fig.* 32. The source remains fixed at L, and the mirror M rotates about an axis passing through L. If M be tilted so as to bring the light from the source on to the photometer head (the direct light from L being screened), then the photometer readings give the candle-powers in different directions round the source in the plane of rotation of M.

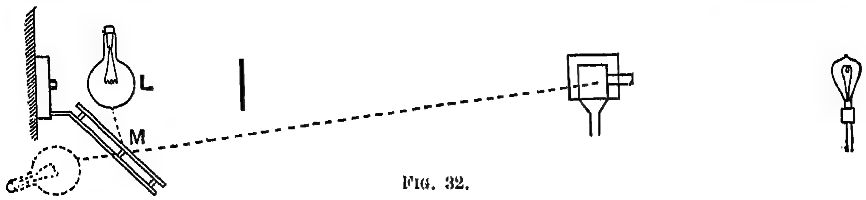


FIG. 32.

(i.) *Errors and Corrections.*—There is one source of error in this apparatus. If the radius about which M swings be not small in comparison with the distance of M from the photometer, then the light is incident obliquely on the photometer disc, so that a change in its reflection ratio as M rotates may cause a small error. Of course, the distance to be used in computing the candle-power must be that of the photometer from the image of the source as seen in the mirror. The reflection ratio of the mirror must also be allowed for, as if the mirror only reflects 80 per cent of the incident light, the measured candle-powers must be multiplied by 1·25. The reflection ratio of the mirror may be readily determined by making a candle-power measurement with the centre line of the mirror M in a horizontal plane, and then rotating L until the part which was originally seen in the mirror from the photometer now faces the photometer directly. The ratio of the candle-power in this position (the mirror being screened) to the candle-power previously obtained, gives the factor by which the mirror measurements have to be multiplied in order to obtain true values of candle-power for the curves.

(ii.) *The Three-mirror Apparatus.*—The error referred to in the preceding paragraph is avoided by the use of three mirrors arranged as shown in plan in *Fig.* 33. The principle of the method remains the same, but the light proceeds axially from the third mirror, and the candle-powers measured, after allowing for the reflection ratio of the mirror

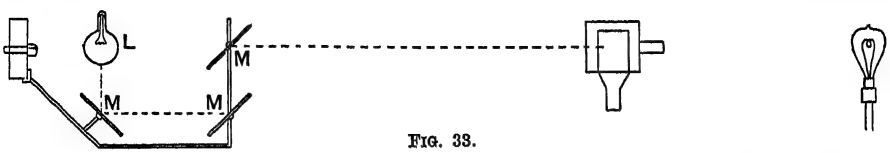


FIG. 33.

[1] Middlekauff and Skogland, *loc. cit.*

system, are not affected by changes in the reflection ratio of the photometer disc.

§ (40) Polar Curves for Large Sources.—The apparatus just described can only be used on the bench for sources of comparatively small dimensions. Very frequently, however, it is necessary to obtain polar curves for sources of large dimensions such as arc lamps, or for a lamp in a reflector fitting where lamp and fitting must be regarded as a single unit for purposes of measurement. In cases such as these, either of the methods described above may be employed, with the necessary modifications in the design of the apparatus. It is, of course, necessary to have the mirrors larger than the source or unit to be measured, and this sometimes necessitates a long swinging arm with consequent accentuation of the error referred to above in the description of the single-mirror method. At the same time, the weight of large mirrors and the difficulty of arranging a three-mirror apparatus for large sources have resulted in the use of the single-mirror method in many laboratories for sources of large dimensions. A polar diagram showing the difference of light distribution from a gas-filled lamp without reflector, and the same lamp with a large shallow reflector, is given in *Fig.* 34.

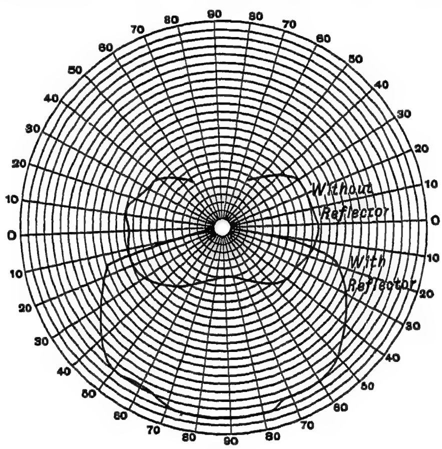


Fig. 34.

With large sources the candle-power is often considerable, and to keep within the best illumination range for accurate photometry it is necessary to have the photometer head at a great distance from the source. This is further desirable when working with a single mirror, as has been pointed out above. It is therefore often convenient to mount the source and the mirror on the wall of the room in which measurements are to be made, and to have a photometer bench of ordinary length mounted on a table on castors so that its distance from the wall can be varied. At the National Physical Laboratory, the bench is actually mounted on steel rollers which run in a track on the floor, and brass marks are fixed close to this track at intervals of a metre, so that, by means of a pointer on one leg of the photometer table, the bench can be moved an accurately known distance away from, or towards, the source.

§ (41) Unsymmetrical Sources.—In all that has been described above, it has been assumed that the source is symmetrical about a vertical axis, so that a polar curve in one vertical plane should be the same for all such planes. This, however, is not the case in practice, and therefore it must be agreed to take the vertical distribution curve in some plane, defined with respect to the source, or the source may be rotated about its vertical axis while the measurements are being made, so that the candle-power shown for any angle θ (measured from the vertical) represents the average value along all the lines forming a cone with the source as apex and semi-vertical angle θ.

When this is done, the speed at which the lamp has to be rotated may be reduced, or the flicker at any given speed may be lessened, by using two mirrors symmetrically placed with respect to the source instead of a single mirror. If this arrangement be adopted it is necessary to cut off the inner corner of each mirror so that both may be used at angles near the vertical.

§ (42) Average Candle-power or M.S.C.P. (i.) *The Rousseau Diagram.*—It might be concluded from the above description that the relative performance of two lamps or lighting units of approximately symmetrical distribution could be seen at once from their respective polar curves, obtained by one of the methods described above; but while it is true that this information can be deduced, after computation, from these curves, the appearance of the curves themselves can be most misleading. This may be very well demonstrated[1] by a consideration of the two curves shown in *Fig.* 35. All the radii vectores of the first curve are double the corresponding ones of the second, so that it is obvious that the total amounts of flux emitted by the two lamps must be in the ratio of two to one. Yet it is equally obvious that the areas of the curves are in the ratio of four to one, while the volumes of their solids of revolution about the vertical axis are in the ratio of eight to one. Clearly neither the area of the polar curve nor the volume of its solid of revolution about the vertical axis can give a mental

[1] Mrs. Ayrton, *The Electric Arc*, p. 454.

conception of the relative amounts of flux emitted by the lamps. This can only be obtained by computation of the average candle-power of the lamp measured in all directions in space. This figure, known as the average candle-power (a.c.p.) or mean spherical candle-power (m.s.c.p.) of a source, can be obtained by calculation from the polar curve in the case of a source whose candle-power performance is symmetrical about the axis of the polar curve. For, if J be the candle-power in a direction making an angle θ with the vertical, then if we suppose a sphere of radius r to surround the source, the area of the zone of this sphere from which the source appears to have the candle-power J is clearly $2\pi r^2 \sin\theta d\theta$, so that the average candle-power is $\frac{1}{2}\int_0^\pi J \sin\theta d\theta$. The value of this expression may be obtained by a simple graphical method due to Rousseau[1] and termed the Rousseau diagram. *Fig.* 36 shows on the left the polar curve of a source of light. At the ends of the radii vectores, horizontal lines are drawn through a vertical line AB, and from the point of intersection of any such horizontal a length is cut off equal

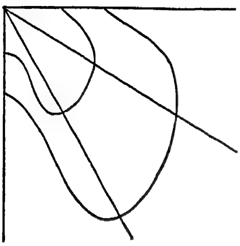


Fig. 35.

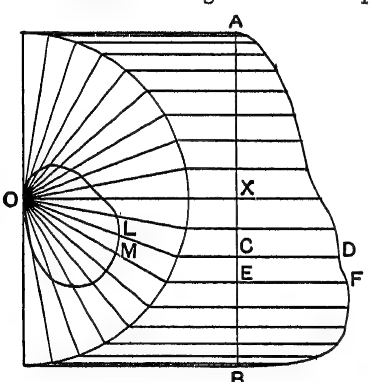


Fig. 36.

to the length of the corresponding radius vector on the polar curve. Thus CD is equal to OL, EF to OM, and so on. A smooth curve is then drawn through all the points such as D, F. From the method of construction of the diagram it will be clear that the distance CX is equal to $r \cos\theta$, so that, in the limit, CE $= r \sin\theta d\theta$, and therefore half the area of the curve ADFB gives the average candle-power of the source. This area may be obtained either by means of a planimeter, or by erecting a series of equidistant ordinates on AB as base, and using one of the forms of Simpson's rule.

(ii.) *Russell Angles.*—In the above method of obtaining the Rousseau diagram it will be seen that measurements of candle-power are made at regular intervals of 10° (or whatever interval may be selected), and that these measurements are then spaced to give them their correct respective weights in determining the area of the curve. Russell's method[2] consists of a predetermination of the angles at which measurements must be made to give equally spaced ordinates on the Rousseau diagram. The spacing of these angles is shown in *Fig.* 37, where it will be seen that the sphere

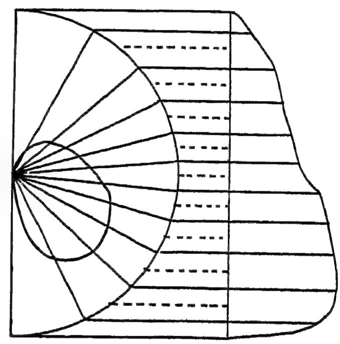


Fig. 37.

is divided by the broken lines into 10 zones of equal area, and then candle-power measurements are made at the half-way points of these zones, so that the area of the Rousseau diagram may be calculated at once by Simpson's rule without any need for the diagram to be drawn. The Russell angles to be used when the sphere is divided into 20, 10, and 5 zones respectively are as follows:

Table of Russell Angles for Calculation of Average Candle-power

20 Zones.	10 Zones.	8 Zones.	6 Zones.
2·9	5·7	7·2	9·6
8·6	17·5	22·0	30·0
14·5	30·0	38·7	56·4
20·5	44·4	61·0	..
26·7	64·2	..	..
33·4	..	..	..
40·5	..	..	..
48·6	..	..	..
58·2	..	..	..
71·8	..	..	..

[1] *La Lumière électrique*, xxxvii. 415.

[2] *Inst. Elect. Eng. Journ.*, 1903, xxxii. 631.

§ (43) BLONDEL'S M.S.C.P. PHOTOMETERS.—The above method of determining the

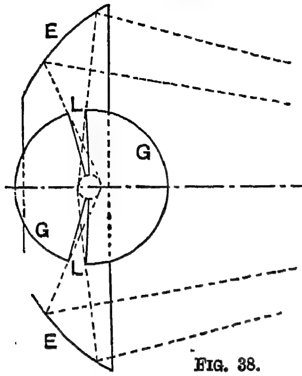


FIG. 38.

average candle-power of a source necessitates the determination of the candle-powers in a certain number of fixed directions. The same information can be obtained in a single measurement by the use of a form of apparatus which concentrates on to the photometer head the light, or a known fraction of it, from a circular ring surrounding the source. Blondel [1] has proposed two different methods of doing this. In one of these, shown in *Fig.* 38, he places the source S at the centre of a spherical blackened globe GG. This globe has cut out of it two lunar apertures LL, and the light from S passes through these and strikes an ellipsoidal mirror E, from which it is brought to a focus at P where the photometer is placed. In a modification of this method the mirror E is replaced by an annular diffusing ring. In each case the instrument is calibrated by means of a lamp of known average candle-power.

§ (44) THE MATTHEWS-DYKE PHOTOMETER.—Matthews' apparatus, as modified by Dyke,[2] is shown in plan and elevation in *Fig.* 39. Eleven pairs of adjustable mirrors M_1M_2 are carried on a semicircular support A. L is the source placed at the centre of the ring of mirrors M_1. These are so placed that the light reaching them from L is reflected to the mirrors M_2 and from the latter proceeds to the photometer head P. The case of this is cut away so that light may reach the photometer screen from all angles. It will be seen that by this method the total illumination of the photometer screen is proportional to the average candle-power of the source L, provided this is symmetrical. For light emitted by the source in a direction making an angle θ with the vertical, on reaching the photometer, makes the same angle with its surface, so that the illumination is reduced by the factor $\sin\theta$. Provided, then, that the diffuse reflection ratio of the photometer screen do not vary with the angle of incidence of the light, the photometer will give the average candle-power directly, if the apparatus be calibrated with a lamp of known average candle-power.

The apparatus was used by Dyke to determine the ratio of the average to the mean horizontal candle-power of a source. For this purpose the light from L is reflected by two other mirrors N, N to the opposite side of the photometer head. These are mounted on a stand which slides on a graduated scale T,

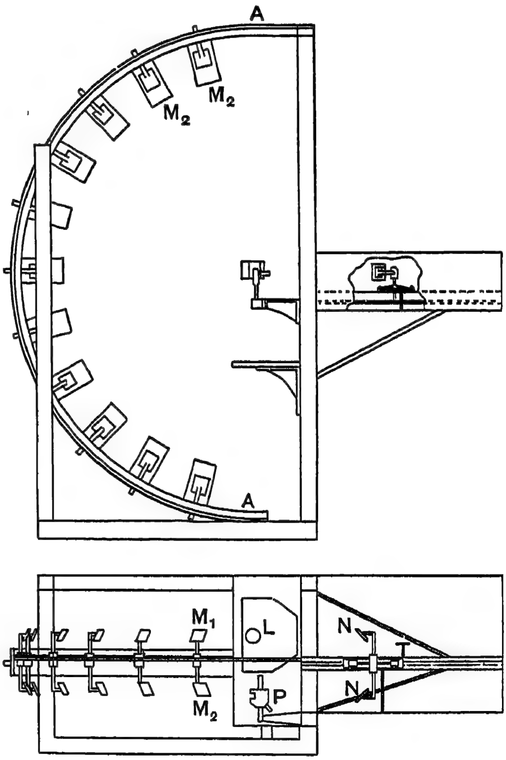


FIG. 39.

and by balancing the photometer the ratio between the mean horizontal and mean spherical candle-powers can be obtained by a single measurement. For the method of correcting for the lack of constancy of reflec-

[1] *Comptes Rendus*, 1895, cxx. 311 and 550.
[2] *Phys. Soc. Proc.* xix. 399, and *Phil. Mag.*, 1905, ix. 136.

tion ratio of the photometer screen, Dyke's original paper should be consulted.

§ (45) INTEGRATING PHOTOMETERS.—In all the above methods of determining the average candle-power it has been necessary to assume that the polar curve is the same in all planes passing through the axis of the lamp, or else that this is sufficiently nearly the case for rotation of the lamp to give a true mean. In the apparatus now to be described this assumption is not made. The distribution of light from the source may be quite irregular, and yet the correct value of average candle-power will be obtained by a single measurement, provided the theoretical conditions of the apparatus be sufficiently closely fulfilled. Actually, as will be seen, the departures from these conditions rendered necessary by practical considerations make the values inexact for very unsymmetrical sources, and the cause of these errors and their elimination will be the subject of the concluding paragraphs of this section.

§ (46) THE WHITENED SPHERE. — W. P. Sumpner first pointed out in 1892[1] that if ABCD (*Fig.* 40) be a principal section of a globe, with a perfectly matt white interior surface, then the amount of radiation reaching any point of the surface B from an element of the surface A was the same whatever the relative positions of A and B. For if O be the centre of the sphere, and F the radiation emitted from A in the direction AO, the radiation emitted in the direction AB will be F cos OAB. Also this radiation strikes the surface at B at an angle from the normal equal to OAB. The amount of radiation reaching B is therefore proportional to F $\cos^2$ OAB/AB^2. But AB $=2r$ cos OAB, and hence this expression becomes $F/4r^2$, which does not depend on the positions of A and B. Hence the radiation received from A by all parts of the interior of the sphere is the same, since any two points of a sphere can be joined by a great circle.

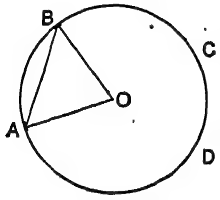


FIG. 40.

If, then, a source of light be placed inside a whitened sphere, a certain amount of light from it reaches each part of the surface of that sphere. The amount diffusely reflected by that part to every other point of the sphere is the same and is proportional to the diffuse reflection ratio of the surface (for normally incident light if the source is placed at the centre of the sphere). Thus any particular spot on the sphere receives, in addition to its own share of the direct light from the source, a constant proportion of the light received by every other point of the sphere, and thus the illumination of a given point shielded from the direct light is proportional to the light received by all the other parts of the sphere, *i.e.* to the average candle-power of the source.

The exact mathematical investigation is as follows: If ϕ be the amount of flux per unit area which reaches the point A from the source, and if R be the reflection ratio of the surface of the sphere, then the amount of flux per unit area which reaches every other part of the sphere, due to reflection from A, is $R\phi/\pi \cdot 1/4r^2$, since ϕ, the flux reflected normally by a perfectly diffusing surface, is equal to the flux received by that surface divided by π (see section on "Illumination"). It is clear, therefore, that if Φ be the total flux emitted by the source, the amount of flux received, per unit area of the sphere, by a single reflection from each other part of the sphere, is $R\Phi/\pi \cdot 1/4r^2$. Similarly, the amount received by two reflections is $R^2\Phi/\pi \cdot 1/4r^2$, and so on. Hence the total flux received by reflection at any point of the sphere is found to be $\Phi/4\pi r^2\{R+R^2+\ldots \text{to infinity}\}=\Phi R/4\pi r^2(1-R)$. But if J be the average candle-power of the source, $\Phi=4\pi J$ (see § (2), "Definitions"), so that the above expression reduces to $J/r^2 \cdot R/(1-R)$. But J/r^2 is the flux per unit area reaching the surface of the sphere, supposing all reflections absent and the source uniform in all directions. In the case of a sphere of one metre radius, a source of one candle would produce an illumination by direct light of 1 metre-candle. If the diffuse reflection ratio of the surface of the sphere be 80 per cent, the illumination by reflected light is $0{\cdot}8/(1-0{\cdot}8)=4$ times as great as this, *i.e.* 4 metre-candles.

§ (47) THE ULBRICHT GLOBE. (i.) *Description.*—The first proposal to use this principle for the determination of average candle-power was made by Ulbricht in 1900,[2] and many developments of the design and contributions to the theory of the sphere photometer have been made by him and others since that date.[3] Recently a large photometer of this type has been constructed at the Bureau of Standards. This consists of a sphere of 88 inches internal diameter, built up of reinforced concrete on a steel network, and finished off inside to a truly spherical surface.[4] There are two holes in the sphere as shown in the sectional diagram,

[1] *Phys. Soc. Proc.*, 1892, xii. 10.

[2] *Elekt. Zeits.*, 1900, xxi. 595.

[3] Ulbricht, *Elekt. Zeits.*, 1905, xxvi. 512; 1906, xxvii. 50, 803; Bloch, *Elekt. Zeits.*, 1905, xxvi. 1047, 1074; 1906, xxvii. 63; Corsepius, *Elekt. Zeits.*, 1906, xxvii. 468; Monasch, *Elekt. Zeits.*, 1906, xxvii. 669, 695; Bloch, *Illum. Eng.*, London, 1908, i. 274; Corsepius, *Illum. Eng.*, London, 1908, i. 801, 805; Sharp and Millar, *Am. Illum. Eng. Soc. Trans.*, 1908, iii. 502.

[4] Rosa and Taylor, *Am. Illum. Eng. Soc. Trans.*, 1916, xi. 453.

Fig. 41. The top hole is covered with a flat wooden disc which can be lowered from above in annular sections, so that a lamp can be suspended inside the sphere from above if

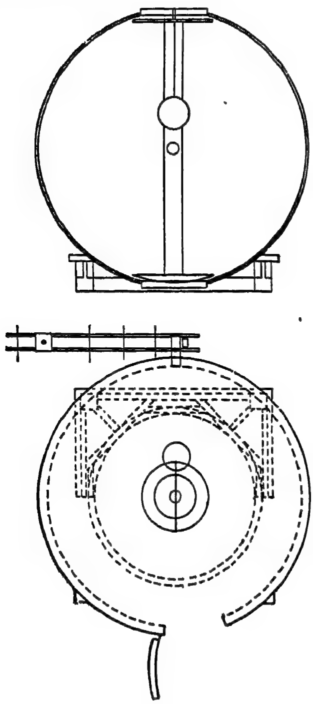


FIG. 41.

desired. On one side of the sphere is a hinged door of segmental form with maximum dimensions $37 \times 16{\cdot}5$ inches. In the wall directly opposite the door, on the equator, is a milk glass window which can be removed at will, but which is perfectly flush with the inside surface of the wall when in place. By an ingenious arrangement of hinged rods carrying the lamp socket, lamps can be brought to the door of the sphere for changing and then automatically returned to their correct position within the sphere. At a point about 27 inches in front of the window are two vertical rods which hold a runner for carrying the screens. These are of four sizes, viz. 11, 21, 30, and 37 cm. in diameter.

As has been stated already, the inside coating of the sphere must be as non-selective as possible, owing to the number of reflections which much of the light has to suffer. In many cases a pure zinc white has been found satisfactory. In the Bureau sphere the inner coating is of Keene's cement, which was found to have a reflection ratio of 92 per cent.

The photometric apparatus consists of a 1·5 metre bench with a photometer head specially designed for the direct comparison of the brightness of the sphere window with the brightness of a diffusing glass illuminated by the comparison lamp.

(ii.) *Errors.*—Tests were made to determine the magnitude of the errors introduced into average candle-power measurements by lack of uniform distribution of light from the source. The maximum error found for many sources having different types of distribution was 1·7 per cent. The percentage reduction of the measured value by the presence of black discs in the sphere was found to be 10 times their relative area (*i.e.* ratio of area of disc to area of sphere surface). For white discs, such as the screens, this reduction was about one-third of that for black discs. Tests with a source giving a beam of light showed a maximum variation, according to the orientation of the beam, of 4 per cent. The effect of the distance of the source from the window was found to be 1 per cent with the lamp half-way between the window and the centre of the sphere. With the lamp at a distance of 10 inches from the window the error was 2 per cent.

(iii.) *Calibration.*—At the Bureau of Standards, the sphere is calibrated before each period of use, by means of a lamp of accurately known average candle-power inserted in the same lamp socket as that which subsequently holds the test lamps. This is the method also adopted with the cube photometer used at the National Physical Laboratory, to be described later.

Ulbricht's original suggestion, however, was to have both test lamp and sub-standard in the sphere during both calibration and test. This arrangement is shown in *Fig.* 42, where L_1 is the sub-standard, L_2 the test lamp, and S_1

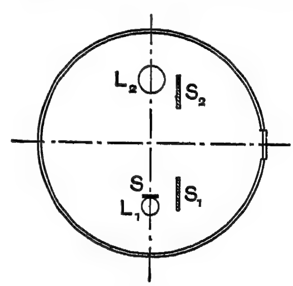


FIG. 42.

and S_2 screens shielding the window from these two sources respectively. A small screen S is also inserted to prevent specular reflection from one lamp when the other is alight. Calibration is effected by balancing the comparison lamp with L_1 on and L_2 off. The tests are then made with L_1 off and L_2 on.

If the window of the sphere does not form part of the photometer head, the two must be kept rigidly fixed in relation to each other. A useful adjunct is an iris diaphragm between the two, so that, when using the sphere for sources of very high candle-powers, the illumination on the photometer can be reduced. Of course, the diaphragm must be at the same aperture for both calibration and test.

(iv.) *Corrections.*—It will be seen from the theory outlined above that the illumination of the window is only truly proportional to the average candle-power of the source so long as the sphere is perfectly empty. The very presence in the sphere of a source of finite dimensions is a violation of this condition, and the fact that screens have to be introduced to shield the window from direct light at once introduces a further departure from the ideal conditions. The error caused by the presence of these bodies in the sphere is greater the larger their dimensions compared with those of the sphere. Ulbricht has laid it down that the diameter of the sphere should be not less than six times the diameter of the globe of the largest lamp to be measured in it. The screens, also, must be as small as possible and whitened on both sides. Where Ulbricht's arrangement is adopted the approximate error due to the screens, S and S′, provided these do not exceed 5 per cent of the area of a principal section of the sphere, has been given as $(80s-100s')$ per cent to be added to the measured value, where s and s' are the ratios of the areas of the screens S and S′ to the cross-sectional area of the sphere. Where the screens are larger than this, more complicated formulae must be employed, and for these Ulbricht's original paper should be consulted (*loc. cit.*). Where the substitution method of calibration is used, the error due to the screen may be taken as that found experimentally and given above under the description of the Bureau of Standards sphere.

§ (48) The Whitened Cube.—Various modifications of the Ulbricht globe have been suggested. One of these consists of a whitened hemisphere, another takes the form of a whitened cube. This was suggested by Sumpner [1] and has been adopted at the National Physical Laboratory.[2] Though theoretically less accurate, this form possesses the advantages of simple construction and greater ease of manipulation of the light sources inside it, and for the comparison of lamps of similar light distributions it has been found to give very accurate results. Buckley has shown [3] that for sources as dissimilar in light distribution as a tungsten-filament vacuum lamp, unshaded, and a similar lamp of which the whole of the upper hemisphere has been covered by an opaque shade, the difference introduced into the values of average candle-power measured by the cube do not exceed 4 per cent.

In all types of the globe photometer and its modifications, it is of great importance that the reflection ratio of the diffusing surface shall be uniform throughout; and if a photometer of this type is much used, especially for arc-lamp work, it must be repainted at frequent intervals, as, although frequent calibration will avoid error due to an even deterioration of the whole surface, it will not allow for a greater deterioration in one part of the surface than another.

§ (49) Average Candle-power Measurements in Specified Zones.—For certain purposes it is sometimes more important to measure the total amount of light given by a source in the lower hemisphere. This is expressed by taking the average of the candle-powers of the source measured in all directions below the horizontal plane, and is termed the average candle-power (lower hemisphere) or mean lower hemispherical candle-power. Its value may be obtained by obvious modifications of the methods described above for the determination of average candle-power. Less frequently used values are (i.) the mean upper hemispherical candle-power, and (ii.) the mean zonal candle-power, *i.e.* the average of the candle-powers of the source measured in all directions within a given zone defined in any particular case.

§ (50) General Considerations.—There can be no doubt that the use of an Ulbricht globe, or one of its modifications, will become of increasing importance owing to two facts. First, the tendency in the design of modern illuminants is to pay much more attention to the total light emitted by a source than to the particular manner in which that light is distributed, so that the average candle-power is of much more significance than the candle-power in any arbitrary direction or directions (such as the mean horizontal candle-power). Secondly, the impossibility of rotating many present-day illuminants—even the gas-filled lamp changes its candle-power value when rotated—makes a photometer such as the Blondel or Matthews useless for the determination of the average candle-power. There seems to be little doubt, therefore, that the use of the Ulbricht globe—in comparatively small sizes for the measurement of lamps of the size generally employed in interior lighting—will shortly become universal in photometric laboratories.

A very complete bibliography of the integrating photometer is given by Rosa and Taylor (*loc. cit.*) and by Ulbricht in *Das Kugelphotometer*.

[1] *Illum. Eng.*, London, 1910, iii. 323.
[2] Paterson, Walsh, Taylor, and Barnett, *I.E.E. Trans.*, 1920, lviii. 83.
[3] *Inst. Elect. Eng. Journ.*, 1921, lix. 143.

V. Illumination and its Measurement

The previous sections of this article have been concerned with the measurement of the candle-power of a source of light. In the present section, however, no attention whatever is paid to the source. All that is dealt with is the actual illumination received at a surface, and this may be due to one or many sources of the same or of different kinds.

A very frequent error in terminology is the confusion of "illumination" with "brightness." The distinction has been mentioned in the definition of the latter term in the section on "Definitions," but it may be further elaborated here. Illumination is concerned only with the amount of radiation which reaches a given surface. A piece of white paper and a piece of black velvet lying side by side may well be equally illuminated, but their brightnesses will be very different. The brightness is concerned only with the luminous radiation which a surface emits in the direction of the eye, and therefore two surfaces of different reflection ratios, but equally illuminated, will have brightnesses which are proportional to their respective reflection ratios.

In illumination photometry, as in candle-power photometry, the actual comparison is always between the relative brightnesses of two surfaces, and this necessitates a brief consideration of the way in which the brightness of any given surface varies with the illumination and the angle at which it is viewed.

§ (51) Matt and Polished Surfaces.—Surfaces may be very crudely divided into two classes, polished and matt. In the case of a polished surface, such as a metal mirror, very nearly the whole of the reflected light is emitted in a single direction, viz. that making an angle with the normal equal to the angle which this normal makes with the incident light. A perfectly matt surface, however, returns light according to a cosine rule. This rule states that, for a perfectly diffusing surface, the radiation emitted in a direction making an angle θ with the normal to the surface is equal to $\phi \cos \theta$, where ϕ is the radiation emitted in the normal direction. Thus if A (*Fig.* 43) be an element of a perfectly diffusing surface, the polar curve of distribution of the light emitted from A is the circle ABC, for AC equals AB cos θ. From this rule it results that the total flux emitted by such a surface is

$$\int_0^{\frac{\pi}{2}} \Phi \cos \theta \, . \, 2\pi \sin \theta d\theta \text{ if } \Phi \text{ is the flux emitted}$$

per unit solid angle in the normal direction. This expression is equal to $\pi\Phi$, so that the average candle-power of A is $\frac{1}{2}\Phi$ or the mean hemispherical candle-power is $\frac{1}{2}\Phi$.

The brightness of a surface should generally be expressed in terms of candles per unit area, so that a square centimetre of a surface which is emitting Φ lumens per unit solid angle in the normal direction has a brightness of Φ candles per square centimetre in that direction. If the direction be not normal to the surface, the brightness must be expressed in candles per unit area of the surface projected perpendicular to the direction. For a square centimetre viewed in a direction making an angle of 60° with the normal produces an image on the retina which is only half the area of that produced by a square centimetre of surface viewed normally. Hence to obtain the illumination of the image on the retina, the criterion of brightness as far as visual effect is concerned, the radiation emitted per unit projected area must be considered.

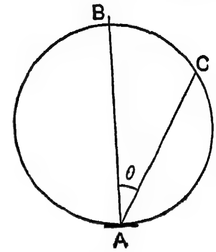


Fig. 43.

It will be seen at once that, since this projected area bears to the true area a ratio equal to the cosine of the angle of projection, it follows that a perfect diffuser appears equally bright at whatever angle it is viewed, for if the brightness be Φ candles per sq. cm. in the normal direction, the radiation emitted per unit solid angle by unit projected area (*i.e.* sec θ sq. cm.) will be Φ sec θ in the normal direction, or Φ in the direction of vision.

It is for this reason that matt surfaces, *i.e.* surfaces which approximate as closely as possible to perfect diffusers, are always employed in illumination photometry. For then the angle from which the comparison surface is viewed is less important than it would be if a glazed or semi-polished surface were employed.

§ (52) Relation between Illumination and Brightness.—It will be useful here to obtain the relation between the illumination of a perfectly diffusing surface, its reflection ratio, and the resulting brightness. If the illumination be E metre-candles, the flux received per unit area of the surface is E lumens. The total amount of flux emitted is therefore RE if R be the reflection ratio. Now if B be the brightness in candles per unit area, the flux emitted per unit solid angle in the normal direction is B lumens. Hence the total amount of flux reflected is πB lumens. Therefore $\pi B = RE$ or $B = RE/\pi$. As an example, a perfectly diffusing surface having a reflection ratio of 75 per cent and illuminated to the extent of 10 metre-candles, has a brightness of $7{\cdot}5/\pi = 2{\cdot}39$ candles per square metre.

§ (53) Photometer Screens.—It has been found that no surface behaves at all angles

as a perfect diffuser, and in some which are generally regarded as matt the departure from the cosine rule is very considerable. This departure depends not only on the direction from which the surface is viewed, but more especially on the direction of the incident light. There is always some specular reflection, and the brightness is always high where the direction of vision and the direction of the incident light make equal angles with the normal in the same plane. Trotter[1] has made a large number of measurements of this effect in the case of many surfaces. He finds that Bristol board, unglazed by passing a damp cloth over its surface, or white celluloid rendered matt by rubbing over its surface with pumice powder, gives very good results if the direction of specular reflection be avoided, a maximum error of 3 per cent being found for the last-named material.

In some photometers the comparison surfaces, or one of them, consist of sheets of matt opal glass, and for these the rule of cosine emission holds very well over quite a wide range of angles.

§ (54) Illumination Photometers.—A general account of the principle of working of illumination photometers was given at the beginning of the section on "Photometric Methods." The number of instruments which have been designed for the measurement of illumination is very large, and all that can be given here is a brief description of some of those which have been most frequently employed by workers in the field of illumination engineering.

§ (55) The Preece Photometer.—The first photometer to be designed for the purpose of making illumination measurements was that of Preece described in 1883.[2] This consisted of a box containing an electric glow-lamp and furnished at the top with a Bunsen grease-spot screen. The current through the lamp was varied until the grease spot disappeared, and the illumination of the screen was then known from previous calibration against the illumination produced by a standard source of light at various known distances. In a modification of this instrument, due to Trotter, the variation in the illumination of the under side of the grease-spot screen was effected by altering the position of the lamp instead of its candle-power.

§ (56) The Trotter Illumination Photometer.—The latest form of the Trotter Illumination Photometer is shown in vertical section in *Fig.* 44. L is a small 4-volt glow-lamp mounted in a screw socket which is carried on a bracket sliding on a vertical bar B. By this means the distance of L from a mirror M can be varied to suit the candle-power of the particular lamp used in the photometer at any time. The light is reflected by M to a matt white celluloid screen C which is capable of rotation about an axis perpendicular to the plane of the paper. This rotation is effected by means of the snail cam A which bears on a pin D. This cam is so shaped that

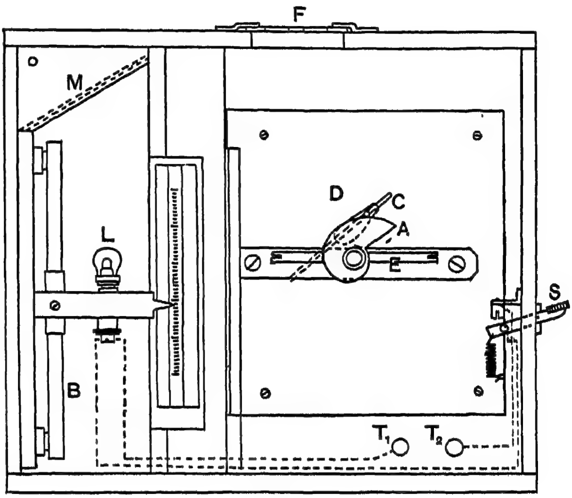


Fig. 44.

the angular motion of the celluloid surface is much slower than that of the cam at the positions where the light from M falls very obliquely on to C. For at these positions it follows from the cosine law that the illumination will vary very rapidly with the inclination of C, so that unless a cam such as that shown is provided, the scale becomes very compressed at the lower values of illumination. A light leaf spring E gives just enough friction to hold the screen in any position while yet allowing a very free movement of the cam. The pin D is held in close contact with the cam by means of a flat spiral spring fixed to C. S is a knife-switch by means of which the lamp L can be lighted from a 2-cell accumulator connected to the terminals T_1, T_2 of the photometer. At the top of the box is a second matt white celluloid surface F, and the photometer is placed so that this surface is in the spot at which it is desired to measure

[1] *Illumination: its Distribution and Measurement*, p. 93, and *Illum. Eng.*, London, 1919, xii. 243.
[2] *Roy. Soc. Proc.*, 1884, xxxvi. 270.

the illumination. A plan view of F is shown in *Fig.* 44A, and the measurement is made by viewing C through the slit in F and adjusting the brightness of the former by tilting it with the cam handle, until F and C appear equally bright. Illumination at F is then given by the position of a pointer connected with the handle by which C is turned. The scale is obtained by previous calibration with known illuminations provided by a standard lamp at different fixed distances. To ensure that F is always viewed at a constant angle, C is provided with two small black pointers. These must be just visible at the ends of the slot in F (as in *Fig.* 44A) when the measurement is being made. The direction actually used is 20° from the vertical. Even with this precaution of constant angle of vision, if the sheet F is not perfectly matt and the light is incident upon it at an angle of about 20°, there is danger that specular reflection may cause an appreciable error. It is best, therefore, to view the photometer in a plane perpendicular to that of the incident light, as shown in *Fig.* 44B, which also gives a view of the complete instrument.

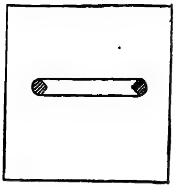


FIG. 44A.

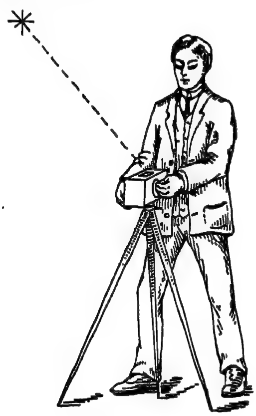


FIG. 44B.

§ (57) THE HARRISON PHOTOMETER.—In the Harrison photometer, the surface illuminated by the outside lights consists of a disc from which two diametrically opposite quadrants have been removed. This disc is set spinning pneumatically, and thus the instrument acts on the flicker principle and is adapted for use with lights of colours different from that of the comparison lamp contained in the photometer box (see section on "Colour Photometry").

§ (58) THE WEBER PHOTOMETER.—The Weber photometer was one of the first illumination photometers designed.[1] It is shown in vertical section in *Fig.* 45. L is a benzine lamp which acts as a source of standard candle-power, S is a translucent screen movable along the tube T_1, its position being marked by a pointer moving over a scale engraved on the outside of this tube. C is a Lummer-Brodhun cube, and P is a total reflection prism for use when the tube T_2, which is capable of rotation about the axis of T_1, is used in the vertical position (as

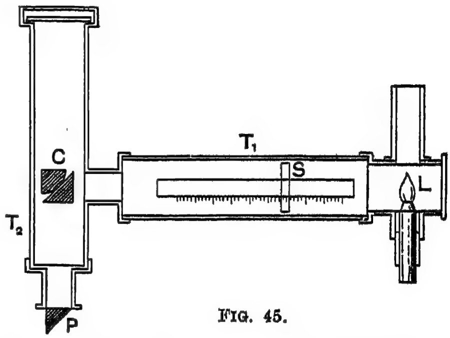


FIG. 45.

shown). The upper end of T_2 is closed by an opal glass disc, and this forms the surface the illumination of which is to be measured.

§ (59) THE SHARP-MILLAR PHOTOMETER.—A much more elaborate photometer on a somewhat similar principle is that of Sharp and Millar, described by the latter in a critical paper on illumination photometers.[2] The plan of this instrument is seen in *Fig.* 46. L is the lamp, a 4-volt glow-lamp, which is capable of movement along the box by means of an endless wire moved by the handle H. The position of the lamp is indicated by the

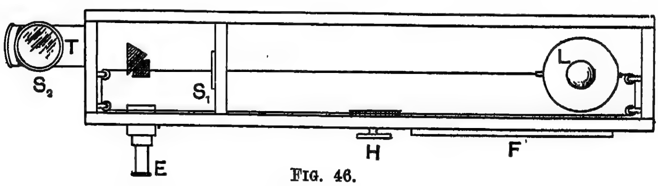


FIG. 46.

[1] *Wied. Ann.*, 1883, xx. 326.
[2] *Am. Illum. Eng. Soc. Trans.*, 1907, ii. 475.

shadow of a pointer on a translucent celluloid scale at F. This lamp illuminates an opal screen S_1, and the brightness of this is compared, by means of a Lummer-Brodhun cube viewed through the eyepiece E, with the brightness of a second opal glass screen S_2 reflected in a 45° mirror, contained in an elbow tube of which T is the plan. The illumination to be measured is that at S_2, and the range of the instrument is increased by the insertion of neutral glasses, with known transmission ratios, between C and either S_1 or S_2.

This photometer possesses the great advantage that the test plate S_2 is viewed from below, so that its illumination is completely unobstructed by the person of the observer or any part of the apparatus. This is of considerable importance, since, when the number of sources contributing to the illumination is large, it is often difficult for the observer to avoid shading one or more of them from the test plate when this is viewed, as is usually the case in illumination photometers, from above.

It is also claimed for this photometer that the brightness of the comparison disc S_1 varies exactly as the inverse square of the distance from it of the lamp L. This is probably the case unless this distance is made too short, when the inevitable effect of interior reflections will be such as to cause departure from the exact inverse square scale.

§ (60) THE MARTENS PHOTOMETER. — The Martens photometer is shown in side elevation in *Fig.* 47 and in section along the line AB in *Fig.* 47A. L is a benzine lamp which is used as a comparison source, and which illuminates a ground glass plate G by

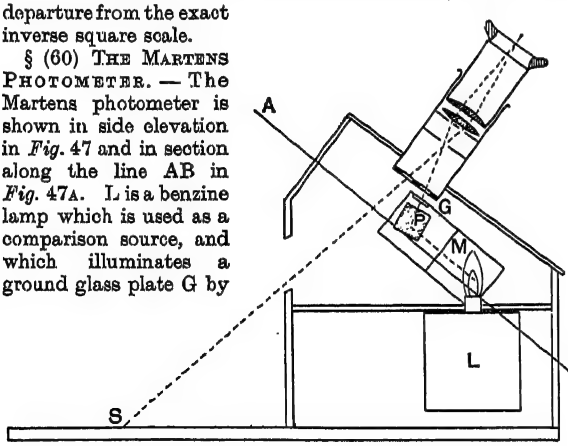


FIG. 47.

reflection in the mirrors M and a total reflection prism P. S is a white screen, the illumination at which is desired, and comparison is made in the eyepiece between the brightnesses of S and G.

A slight modification of the Martens polarisation photometer described in § (30) enables this instrument to be used as an illumination photometer. The plaster screen F is replaced

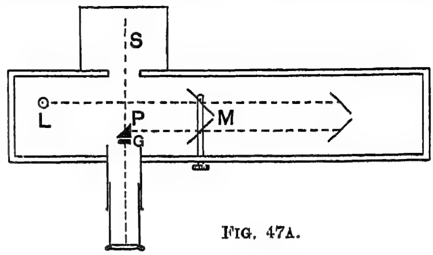


FIG. 47A.

by an opal glass screen, and the tube T is turned so that this screen occupies the position where it is desired to measure the illumination.

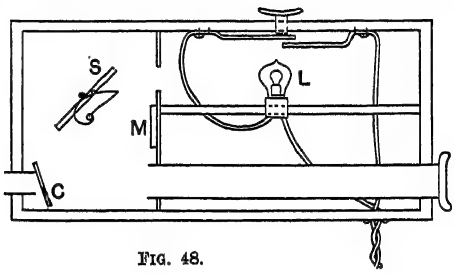


FIG. 48.

§ (61) THE LUXOMETER.—Two very portable instruments employing a test screen which may be quite separate from the photometer are the Luxometer, which is a modification of the Trotter photometer, and the Lumeter due to Dow and Mackinney.[1] The first of these is shown in section in *Fig.* 48. L is a small 2-volt glow-lamp which illuminates the tilting screen S. This screen is reflected in the mirror M and is seen in the plate C which is silvered over one-half of its surface. The test plate T is seen by direct vision through the unsilvered half of C, and the two halves of the field are brought to equality of brightness by tilting S.

§ (62) THE LUMETER.—The lumeter is seen in plan in *Fig.* 49. The 2-volt glow-lamp L illuminates an opal glass plate G. Behind this plate a cylindrical shutter C with a carefully graduated opening is moved by a handle carrying a pointer which travels over a scale on the outside of the instrument,

[1] *Opt. Soc. Trans.*, 1910–11, xii. 66. [This paper describes the original form of the instrument which has now been revived. The description here given refers to a modified form introduced about 1912.]

The light from the part of G exposed by C illuminates the screen S, which consists of an exterior ring of white opaque material, and an inner clear circle. This screen is viewed through the eyepiece, the exterior test plate P being seen through the clear central portion of S, while the brightness of the opaque portion is varied to equality with S by altering the position of C. The shape of the opening in C is seen in the lower part of the figure. The breadth of the larger portion is ten times that of the smaller, so that at the transition point the scale is enlarged ten times. Theoretically the scales should be linear, but in practice errors up to 5 per cent and more are found with a linearly divided scale. Just at the point where the larger opening begins to come into operation, the error is, naturally, very large.

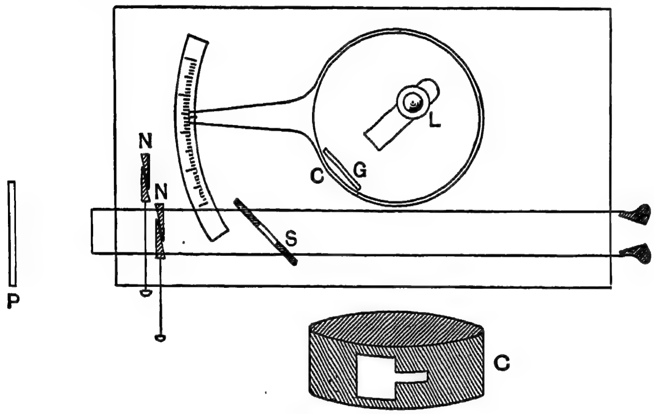


FIG. 49.

Both these instruments are adapted for measuring low illuminations by the provision of neutral glasses N, N, of known transmission ratios, which can be inserted at will between the test surface and the eyepiece. In the lumeter these consist of double wedges, so that adjustment to convenient ratios, such as 0·1 and 0·01, is possible.

§ (63) THE SHARP FOOT-CANDLE METER.—For rapid approximate measurement of illumination a portable photometer has been devised by Sharp and modified by Sackwitz.[1]

It consists of a box containing a trough, over which lies a white screen with a line of translucent circles in it, this screen being illuminated from beneath by a lamp placed at the end of the trough. The brightness of the circles diminishes gradually from one end of the screen to the other, and when the box is placed in any position, one of the circles is found to disappear, those on the right still showing bright, while those on the left now appear dark. Previous calibration gives the value of illumination at which each circle vanishes in this way, and thus the instrument can be used for obtaining quickly an approximate value of the illumination at any point.

The lamp is supplied from a dry cell contained in the box, and a small voltmeter and regulating resistance are supplied to enable the fall of voltage of the cell to be compensated.

A somewhat similar instrument was described[2] by H. T. Harrison in 1910.

§ (64) THE MEASUREMENT OF BRIGHTNESS.—In both the luxometer and the lumeter, the test surface is quite separate from the photometer, and in fact these instruments may be used for measuring brightness as well as illumination. For, if the reflection ratio of the white card be R, then an illumination of n metre-candles gives it a brightness of nR/π candles per square metre. Thus if the balance be obtained by means of the instrument when the test card is replaced by some other surface, the reading of the instrument multiplied by the factor R/π will give the brightness of the surface in candles per square metre. Similarly, if the instrument be calibrated (as is usually the case) in foot-candles, multiplication by R/π gives the brightness in candles per square foot.

It has frequently been the custom to express brightness in equivalent foot-candles, *i.e.* in terms of the brightness of a perfectly diffusing surface of 100 per cent reflection ratio, illuminated to the extent of one foot-candle. This figure is obtained at once from the readings of an illumination photometer multiplied by the simple factor R. Unfortunately this system has given rise to much confusion between illumination and brightness, and it is desirable that the latter should always be expressed in candles per unit area.

§ (65) VOLTAGE REGULATION.—It will have been noticed that a number of illumination photometers depend, for their standard of comparison, on a small 2- or 4-volt electric glow-lamp fed from a portable battery. Now, as has been already stated, the candle-power of an electric lamp varies as the fourth or fifth power of the voltage, so that constancy

[1] Sharp, *Elect. World*, 1916, lxviii. 569; Sackwitz, *Am. Illum. Eng. Soc. Trans.*, 1918, xiii. 292.

[2] *Illuminating Engineer*, London, 1910, iii. 373.

of battery voltage is of first importance. Most such photometers are provided with a switch so that the lamp can be switched on only when the readings are being taken. The discharge of the battery is thereby much reduced, and, as the current of the lamp does not generally exceed half an ampere, a storage cell of 10 to 20 ampere-hour capacity will maintain a constant supply voltage over a considerable period of use. The voltage of the battery on discharge must be frequently checked, and recharging should be commenced as soon as the voltage has dropped 5 per cent of its value. During this time the readings of the photometer must be reduced by 3·7 or 5 per cent for every 1 per cent drop in voltage, according as the lamp filament is of tungsten or carbon. The cell when first taken off charge should be discharged at about half an ampere for at least an hour before being used on the photometer. This avoids the initial over-voltage.

§ (66) Precautions in using a Portable Photometer.—The most frequent source of trouble in portable photometers employing an electric lamp, is faulty contact at some part of the circuit. At the very low voltage used, the slightest fault in a contact causes a noticeable decrease or fluctuation in the light. The leads to the battery should be tightly screwed down on to perfectly clean terminals, and it is inadvisable to undo them during the taking of a set of readings. All contacts inside the photometer should be soldered, and the lamp socket must be of the screw type with the lamp well screwed down into the socket. If a switch is provided for the lamp, it is necessary to ensure that it makes good and constant contact when in the "on" position.

All portable photometers require frequent checking at two or three points of their scale, against known illuminations provided by a standard lamp at definite distances from the test plate. When all the precautions detailed above have been observed, the best of these instruments may be relied upon to an accuracy of about 2 to 3 per cent over the most favourable part of its scale.

The measurement of illumination, both by daylight and artificial light, is becoming increasingly frequent and important. In several countries codes of illumination required for particular kinds of workshop and factory processes are enforced by legislation. Extensive rules have also been framed for the lighting of schoolrooms, and the illumination of shop windows, public halls, churches, theatres, and other buildings has lately received considerable attention.[1]

§ (67) The Measurement of Illumination.—The method of making illumination measurements is to place the test plate at the position where the illumination is required, and to determine this by means of a photometer similar to one of those above described, taking care that the body of the observer shields as little light as possible from the test plate. Unless otherwise stated or clearly implied, it is usual to assume that the test plate is placed horizontally, and very frequently the floor level or the 1-metre level is adopted for all the measurements. More frequently, however, the plate is placed horizontally on the desk, loom, bench, lathe, etc., where it is desired to know the illumination. Sometimes, as in planning a picture gallery, the illumination of a vertical surface is of primary importance. In such cases, of course, the vertical position is adopted for the test plate.

§ (68) The Calculation of Illumination.—If it be desired to predict by calculation the distribution of illumination which will result from any given arrangement of light sources, this may be done from the polar curves of the sources to be used. For (*Fig.* 50) the illumination of a horizontal plane at A, due to a source at L of which the candle-power in the direction LA is J, will be $J\cos\theta/LA^2$ or $J\cos^3\theta/h^2$, where h is the vertical height of L above the horizontal plane through A. Thus the total illumination at A will be $\Sigma J\cos^3\theta/h^2$, the summation being made for all the sources which send light directly towards A.

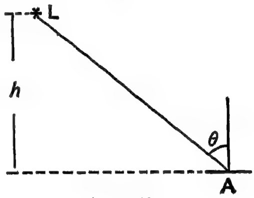


Fig. 50.

In illumination calculations it is usual to assume some simple form of polar distribution which approximates to that of the actual light source to be used. The great variety of the polar curves of present-day sources make it useless to do more than give a single example of such calculations. For other examples the reader is referred to some book on illumination engineering, such as Gaster and Dow, *Modern Illuminants and Illuminating Engineering*; Trotter, *Illumination: its Distribution and Measurement*; Bohle, *Electrical Photometry and Illumination*; Uppenborn, *Lehrbuch der Photometrie*; Steinmetz, *Radiation, Light and Illumination*; etc.

§ (69) The Illumination Curve.—If we have a number of equal light sources of known polar distribution placed in line at a given height, the curve showing the variation of illumination along the line vertically beneath the sources can be very readily obtained. From the expression given

[1] Vide *Illuminating Engineer*, London; *Transactions of the Illuminating Engineering Society of New York*; *Lighting Journal*, and other publications.

above, it is easy to obtain, from the polar curve of the source, the diagram connecting the illumination at a point P, due to a single source, with the horizontal distance of P from the source. This has been done in *Fig.* 51, where the heavy line is the curve of

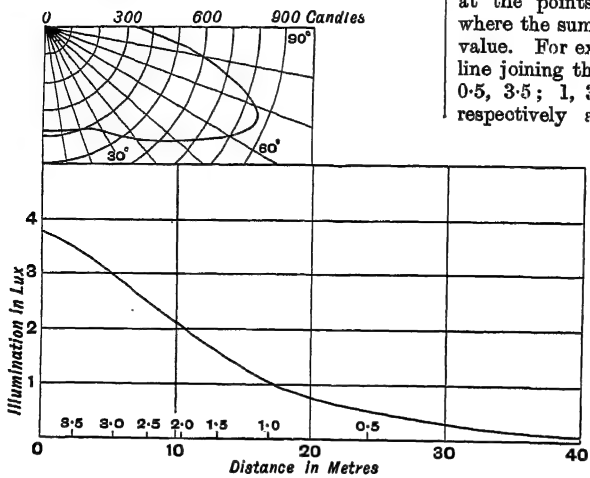


FIG. 51.

illumination along a line 10 metres below a source having the polar diagram shown. If now the sources be spaced 20 metres apart as shown in *Fig.* 52, a curve like that in *Fig.* 51 is placed in its proper position under each source and the ordinates are added to produce the full line curve which shows the distribution of illumination due to all the sources together. It will be obvious that the method may be applied to other arrangements of sources or to a variety of different sources, though the calculations will not be quite so simple as in the case of the example given above.

The curve of *Fig.* 52 is termed an illumination curve. It shows only the distribution of illumination along a single line. If it be desired to show the distribution over an area, this is most conveniently done by means of an iso-lux diagram, or contour map of equal illumination. Trotter[1] has described the method of constructing such a map. The simplest case is that of two sources. Strips of paper are marked off with a scale representing the illuminations due to a single source at various horizontal distances from that source. Such a scale corresponding to the curve of *Fig.* 51 is shown at the base of that figure. Two such scales are pinned to a sheet at points representing the positions of the two sources (see *Fig.* 53) and marks are made at the points of intersection of the scales where the sum of the graduations has a given value. For example, the 4-lux contour is the line joining the points where the scale marks 0·5, 3·5; 1, 3; 1·5, 2·5 on the two scales respectively are coincident. The complete contour map is shown in *Fig.* 53. The larger the number of sources, the more complicated become the necessary calculations. A large number of such maps are given by Trotter (*loc. cit.*), Blondel, Uppenborn,[2] Steinmetz,[3] and Maréchal.[4]

It will be seen that, although either an illumination curve or an iso-lux diagram may be used to obtain a mental conception of the effect of a given system of light sources of known distribution, yet it is still necessary to have some definite figure by which the illumination performances of two lighting systems may be compared. This is analogous to the comparison of the light-giving power of two sources by means of their average candle-powers instead of by the more detailed but comparatively cumbrous polar diagrams.

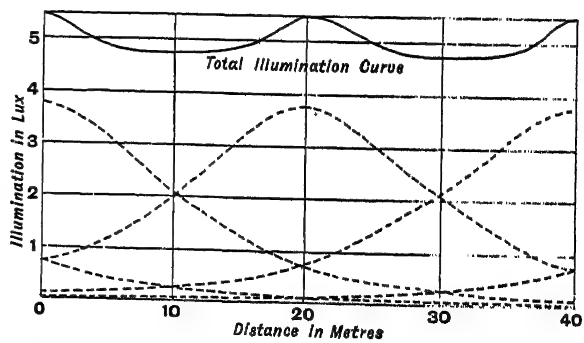


FIG. 52.

§ (70) THE AVERAGE ILLUMINATION. — Various suggestions have been made for a basis of comparison of different illumination systems. On the whole the average illumination, or, what is equivalent to it, the average

[1] *Illumination: its Distribution and Measurement*, p. 46.

[2] Uppenborn and Monasch, *loc. cit.*

[3] Steinmetz, *loc. cit.*

[4] *L'Éclairage à Paris.*

flux per unit area (generally expressed in lumens per square metre or per square foot), has been most generally adopted for indoor

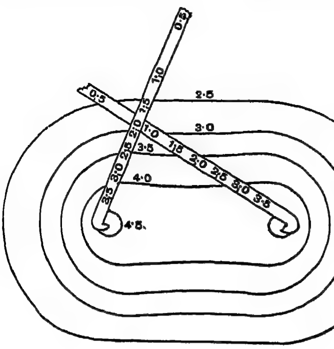


FIG. 53.

illumination. This means that, for a direct lighting system where the reflection from walls and ceiling can be neglected to the approximation desired, the average illumination in metre-candles will be equal to the total mean lower hemispherical candle-powers of all the sources in the room, multiplied by 2π and divided by the area of the room in square metres.

(i.) *The Variation Factor.*—This system gives no indication as to the distribution of illumination over the room. If there be only a few sources of high candle-power the illumination will be concentrated in the regions underneath a source, leaving the outer parts of the room in comparative darkness. For this reason a second figure, giving the ratio of maximum to minimum illumination in the useful area of the room, has been proposed as an addition to the average illumination. This figure may be termed the "variation factor."

(ii.) *Trotter's "Characteristic Curve."*—Another method of describing an illumination system has been called by Trotter[1] a characteristic curve. This is a curve having for its abscissae areas, and for ordinates the values of the minimum illumination over those areas. Thus, referring once more to the illumination diagram shown in *Fig.* 51, it will be seen that the illumination at all points within a circular area of radius 10 metres equals or exceeds 2·1 metre-candles. The area of this circle is 314 square metres, and therefore the point (314, 2·1) is a point on the characteristic curve which is shown in *Fig.* 54. Since the abscissae of this curve represent areas, the mean ordinate of this diagram, as far as any given ordinate, is equal to the illumination over the circle having an area represented by that ordinate. Thus the average illumination over a circle 1000 sq. metres in area is equal to the mean ordinate of the area OABC on the diagram, *i.e.* to 1·86 metre-candles.

(iii.) *Bloch's Method of calculating Average Illumination.*—Two methods which have been proposed for the approximate calculation of the average outdoor illumination due to lamps having a given distribution will be shortly described. In the first, due to Bloch,[2] the area is divided into a number of rectangles so that there is a lamp in the centre of each rectangle. The radius of the circle which has an area equal to that of the rectangle is then found, and the average illumination over this circle is obtained from the polar diagram of the source. This is equal to the total radiation emitted by the source within the cone having

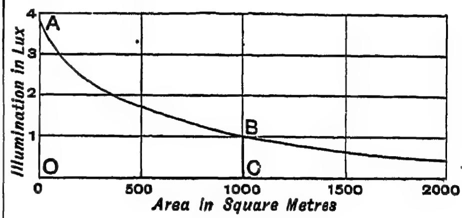


FIG. 54.

the lamp as apex and the circle as base, divided by the area of the circle.

In this method the effect of neighbouring lamps is neglected, making the calculated value too low; but, on the other hand, the conversion of the rectangle into a circle makes the value too high, so that these errors tend to neutralise each other. Bloch proposes that for street lighting the value so obtained shall be multiplied by the factor $(1{\cdot}2 - 0{\cdot}1\gamma)$, where γ is the ratio of the distance between the lamps to the width of the street.

(iv.) *Högner's Method of calculating Average Illumination.*—Högner's method is more accurate, but rather more laborious. The rectangular area surrounding a source is divided into four smaller rectangles with their common corner at the point vertically below the lamp. Each of these four rectangles is then further subdivided into still smaller areas by the construction shown in *Fig.* 55. Points X_1, X_2 . . . Y_1, Y_2 . . . are marked off on AB and AC, so that the lines X_1L, X_2L . . . and Y_1L, Y_2L . . . make angles of 10°, 20° . . . with the vertical. Mutually perpendicular ordinates drawn through X_1, X_2 . . . and Y_1, Y_2 . . . divide the area into a number of small rectangles such as PQRS. Now each of these rectangles is regarded as forming the base of a pyramid having L as apex, and the solid angle ω, and the inclination to the vertical θ, of the line joining its centre

[1] *Loc. cit.* p. 58.

[2] *Elektrot. Zeits.*, 1906, xxvii. 493.

to L has been found and tabulated for each such rectangle. It only remains, therefore, to find the candle-power of the lamp at any angle θ, and to multiply this by the value of ω (in steradians), to find the total flux reaching the base of the pyramid. The total flux over the whole area being thus found by addition, the average illumination can be calculated.

It will be clear that the illumination due to a given system of sources cannot be altered in distribution if the candle-power of all the sources be altered in the same ratio. It therefore follows that to determine the candle-power required for a given illumination E_1, using sources of known distribution arranged in a certain way, it is only necessary to calculate the illumination due to such a system of sources of candle-power J. Suppose this illumination is found to be E_2. Then the candle-power required is JE_2/E_1. This refers equally to illumination at a point and to average illumination over an area.

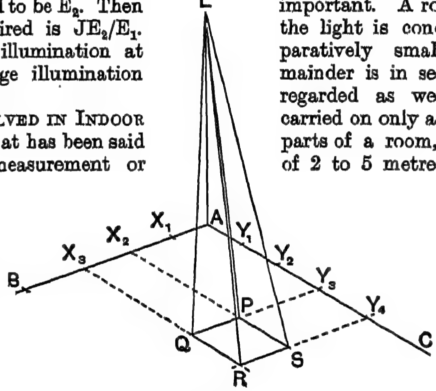


FIG. 55.

§ (71) FACTORS INVOLVED IN INDOOR LIGHTING DESIGN.—What has been said above refers to the measurement or calculation of the actual illumination at a point, or the average illumination over an area, due to a given system of light sources. But it will be obvious that in designing any system of lighting the degree of illumination produced is only one out of several factors, all of which have to receive consideration. Every particular case will present its own special problems, and only the most general rules can be given here as an indication of the principal factors which enter into the design of all lighting systems.

The problems of indoor illumination must be treated quite separately from those of street lighting and outside illumination generally. The requirements of an indoor system may be briefly summarised as (1) adequate intensity, (2) proper distribution, (3) absence of glare, (4) avoidance of excessive contrasts—and these will be dealt with in turn. For a discussion of the particular problems presented by the many classes of indoor illumination, and of the manner in which these may be dealt with, the reader is referred to chapter ix. of *Modern Illuminants and Illuminating Engineering*, by Gaster and Dow.

§ (72) ILLUMINATION REQUIRED FOR VARIOUS PROCESSES. (i.) *The Amount.*—The necessary illumination for various types of room, workshop, etc., depends almost entirely on the nature of the work. For reading, writing, and similar work an illumination of 30 to 60 metre-candles has been proposed. For such places as foundries, where the illumination has to be general and no fine work is done, 10 to 20 metre-candles is generally regarded as sufficient, while for a drawing-office 80 metre-candles is needed, and in the special case of a shop window, where display is required and the goods to be shown are of a dark colour, as much as 200 metre-candles may be used with advantage. Tables of the intensities suggested for various purposes have been frequently put forward.[1] It has been found that for similar work with materials of different reflection ratios the illumination required is inversely proportional to the reflection ratio.[2]

(ii.) *Distribution of Illumination.* — The proper distribution of the light is most important. A room in which nearly all the light is concentrated in one comparatively small area, while the remainder is in semi-darkness, cannot be regarded as well lighted. If work is carried on only at one or more particular parts of a room, a general illumination of 2 to 5 metre-candles should be provided, with the addition of local lights to give the necessary illumination at the points where it is specially needed. For reading, the light should preferably come from behind, while for writing it should come from the left front.

(iii.) *Avoidance of Glare.* — The avoidance of glare is of the utmost importance if eye-strain is to be avoided and the light provided is to be properly appreciated by the eye. Even sources of low intrinsic brilliancy should be shielded from direct vision, and it has been generally agreed that an intrinsic brilliancy of 3 to 5 candles per square inch should not be exceeded in any source, shade, or reflector which the eye is liable to see by direct vision. It is a matter of common experience that exposure to a very brightly illuminated surface renders the eye less capable of properly appreciating the forms of objects having a lower intrinsic brilliancy, so that if glare be not avoided a much higher general illumination is required.

Excessive contrasts, generally in the form of deep shadows, are annoying to the eye in domestic lighting, and may be sources of the utmost danger in street or factory lighting.

[1] Gaster and Dow, *loc. cit.* p. 329.
[2] *Report of Departmental Committee on Lighting in Factories and Workshops*, i. 37.

On the other hand, an almost complete absence of shadow is most unsuitable for work requiring perception of fine detail, such as sewing, watchmaking, etc.[1] The attainment of the above requirements in any particular case is a matter requiring much experience of illumination engineering.

(iv.) *Direct and Indirect Lighting.*—For the sake of convenience, systems of indoor lighting have been classified as direct, indirect, and semi-indirect. In the first-named system, the light from the source reaches the area to be illuminated by a direct path and does not suffer any change of direction once it has left the lighting unit. In an indirect system, on the other hand, the light is all cast upwards to the ceiling, or to a reflector placed above the source, so that the light reaching the lower part of the room has suffered at least one reflection on its way. In the semi-indirect system of lighting, part of the light from the source illuminates the room directly, the remainder by reflection. In addition to the ceiling, the upper part of the walls has a considerable share in the distribution of the light from an indirect or semi-indirect system. Both of these systems, owing to the great area from which the light is received by the lower part of the room, avoid all possibility of glare, and cause a very great softening of shadows. With a totally indirect system, the lack of shadow may be so pronounced as to cause inconvenience if the room is used for certain kinds of work.

Combination of indirect general lighting with direct local lighting is frequent and possesses many of the advantages of both systems. With indirect or semi-indirect systems and, to a less extent, with direct systems of lighting the reflection ratio of walls and ceilings has a marked effect on the intensity of the sources required to produce a given illumination. Walls with a very high reflection ratio may increase the illumination by 40 to 50 per cent over that obtainable with very dark walls.[2] In general it may be assumed that with a white ceiling the sources used in an indirect system of lighting require to have double the intensity of those used for a direct system. Useful tables of the consumption of power by various illuminants, when producing an illumination of one foot-candle on the working plane, have been given by Gaster and Dow [3] and by Uppenborn.[4]

§ (73) Outdoor Illumination.—The requirements in the case of outdoor illumination are somewhat different. In street lighting the need is for as even an illumination as possible with the comparatively wide spacing necessary. This is partly accomplished, either by specially designed reflectors which concentrate the light in a direction only slightly below the horizontal, or by placing the lighting units at a considerable distance above the roadway. A *minimum* illumination of 1 metre-candle is regarded as necessary for important thoroughfares.[5]

Other problems which come under the head of outdoor lighting are the illumination of railway platforms and goods-yards, parks, bridges, and open spaces. The illumination of signboards and building exteriors for spectacular purposes are also cases requiring special treatment. For these a work on illumination engineering should be consulted.

§ (74) Daylight.—What has been said so far has been with reference to illumination by artificial light. But it will be obvious that the methods of illumination measurement, and the photometers which have been described, may be employed equally well for the measurement of daylight illumination. The chief difficulty met with, that of the great colour difference between daylight and the light from the comparison source used in a photometer, may be overcome to a certain extent by the use of a blue glass or gelatine filter between the comparison lamp and the surface it illuminates (see § (101) in "Colour Photometry"). This has the disadvantage of reducing the illumination of the comparison surface, whereas the illumination to be measured is generally high in the case of daylight. This has led to the use of a yellow filter placed between the test surface and the eyepiece, but this has the disadvantage that the transmission ratio of the filter may vary with the colour of the daylight to be measured.

(i.) *Daylight Factor.*—A further complication in the case of measurements of interior illumination by daylight arises from the fact that the daylight is constantly varying in intensity. A simple measurement of the illumination at a point inside a room on a particular occasion, therefore, does not give the information required, for clearly this will depend on the illumination prevailing outside the building, and hence a knowledge of simultaneous values of indoor and outdoor illuminations is essential if any information of permanent value is to be obtained. On a dull day with a grey sky, the illumination at a point inside a building bears a very nearly constant ratio to the illumination at a point outside. This ratio, expressed as a percentage, has been termed the daylight factor at the point and is taken as giving some indication of the illumination at the point when the outside illumination is known.

(ii.) *Variations of Daylight.*—It is seldom

[1] *Departmental Committee on Lighting in Factories and Workshops, loc. cit.*
[2] Gaster and Dow, *loc. cit.* p. 326.
[3] *Loc. cit.* p. 323.
[4] *Lehrbuch der Photometrie*, p. 179.
[5] Gaster and Dow, *loc. cit.* p. 419.

realised how rapid and how great are the variations which take place in daylight illumination. Some experiments on these have been described by C. C. Paterson and the author.[1]

The conditions exerting the greatest influence on daylight illumination are three, viz. (i.) time of day, (ii.) time of year, (iii.) meteorological conditions.

(iii.) *Variation with Time of Day.*—As to the first, experiment has shown that the gradual rise of illumination at sunrise closely corresponds with the fall at sunset, and that the change of illumination takes place at a more uniform rate when there is a dull grey sky than when the sky is blue and cloudless. The sunset and sunrise illuminations do not vary with the time of year, but only with the meteorological conditions, and may be assumed to lie, generally, between 100 and 500 metre-candles; 250 metre-candles may be taken as a fair average value.

(iv.) *Variation with Time of Year.*—As to the variation with time of year, the average results of a very large number of observations extending over nine months gave the curves shown in *Fig.* 56 for the variation of

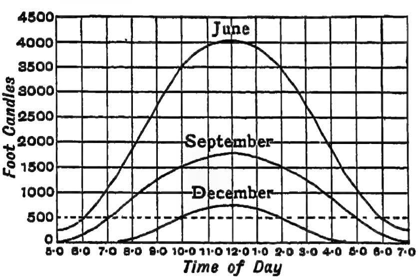


FIG. 56.

daylight from 5 A.M. to 7 P.M. on average days at different times of the year. On this diagram the horizontal scale represents the time of day, and the vertical scale foot-candles of illumination under an open sky. There are three curves, representing average days in June, September, and December respectively. A horizontal line has been drawn through the curves at the 500 foot-candle level for the purpose of illustration, and it will be seen that the line cuts the December curve at 10 A.M. and 2 P.M. It follows that any room in a building for which the daylight factor is below 0·2 per cent will have less than 1 foot-candle of daylight illumination before 10 A.M. and after 2 P.M. on an average day in December, while in June the corresponding times are 6 A.M. and 6 P.M. respectively.

It may be said, further, that there is, on the average, an outside illumination of about 100 foot-candles 30 minutes after sunrise and before sunset, so that a position in any building with 1 per cent daylight factor has, on the average, an illumination of 1 foot-candle 30 minutes after sunrise and before sunset.

Although the sunrise and sunset illuminations do not vary appreciably from summer to winter, the mid-day illumination falls from an average of about 4000 foot-candles in June to 700 foot-candles in December, the intermediate distribution being of the nature of that shown in *Fig.* 57.

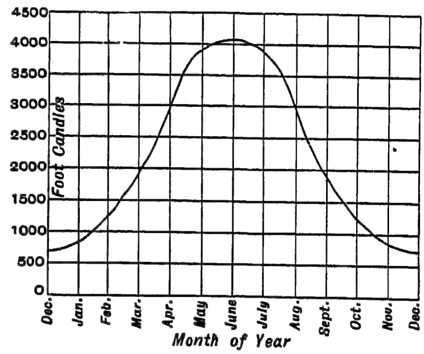


FIG. 57.

(v.) *Variation with Meteorological Conditions.*—The variations in the daylight illumination arising from even small changes in atmospheric or cloud conditions are often surprisingly great and take place with extreme rapidity. The diagrams in *Fig.* 58 show the magnitude of

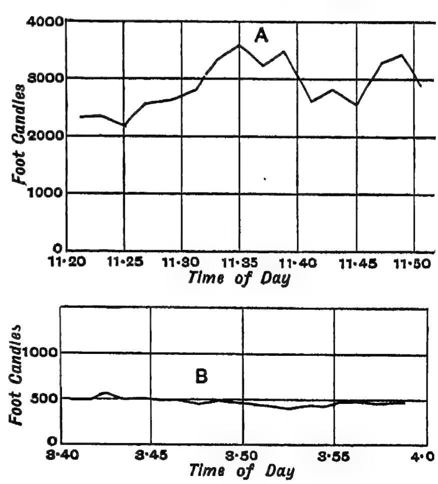


FIG. 58.

such variations on representative days. The first gives an approximate record of the day-

[1] *First Report of the Departmental Committee on Lighting in Factories and Workshops*, 1915, i. 63.

light illumination observed on a bright day in March (17th), when there was a large number of grey and white clouds over the sky. The second shows the comparative steadiness found on a day with a very uniform and dull grey sky at the same time of year (March 30th).

It will be seen from the first example that the illumination may vary by upwards of 80 per cent within a few minutes. Also two days of a different type, though in the same month, may give widely different average illumination values at the same time of day. Thus the midday illumination on one day when it was very dull and drizzling was 150 foot-candles, while three days later, at the same time of day, it was 1200 foot-candles. Of the various factors affecting the illumination, the number and brightness of the clouds are, perhaps, the most important. A bright cloud upon which the sun is shining has a brightness far exceeding that of the surrounding blue sky, and as such a cloud changes its position in the heavens its brightness, as well as its capacity to produce illumination at any particular position, will both change considerably. With a large number of such clouds in the sky it is not surprising that the illumination should fluctuate rapidly. The clearness of the atmosphere, too, has naturally a considerable influence on the illumination, but this effect is not capable of such rapid variations as that due to clouds. The probable limits between which the daylight illumination at any time may be expected to lie are indicated in the table. The values there given are based on 9½ months' observations taken at the National Physical Laboratory from March to December 1914. Observations over a more extended period might lead to some modification of the averages given. The values of illumination considered throughout are those measured on a horizontal plane placed in an open field with a practically unobstructed horizon. The direct sunlight was always shielded from the test cards.

TABLE

Month.	Midday Illumination (Foot-candles).		
	Highest.	Lowest.	Mean.
1914.			
April	5340	710	3200
May	5430	1050	3200
June	5160	2420	4100
July, August	Sufficient observations not available.		
September	3030	870	1180
October	2610	152	1300
November	2500	130	1100
December	1160	360	720

§ (75) MEASUREMENT OF DAYLIGHT FACTOR.—In view of what has been said above as to the variability of daylight, it would appear that the only satisfactory method of specifying the daylight illumination of a room is by its daylight factor, either the minimum, or an average for the whole room, being taken. This involves simultaneous measurements of illumination inside and outside the building.

A convenient method of arranging this is to have one observer taking readings at regular one-minute intervals on a test plate so situated that it receives unobstructed light from the whole sky. At the same time, another observer inside the building takes readings at convenient intervals (consisting of an integral number of minutes), and the daylight factors are subsequently calculated from the two sets of data. This was the method adopted by the author in making 4000 measurements of daylight factors for the Home Office Committee on Lighting in Factories and Workshops. (*Report*, vol. iii.)

§ (76) TROTTER PHOTOMETER DAYLIGHT ATTACHMENT.—There is one difficulty in applying the method to buildings in crowded areas, viz. that of finding a sufficiently open space for the outside measurements. This may be partially overcome by the adoption of a device described by Trotter and Waldram. It may be readily shown that if E be the illumination of a test plate which is receiving unobstructed light from a complete hemisphere of sky of brightness B, then $E = \pi B$. If now a vertical tube be placed over the test plate, and if an opening at the upper end of this tube be of such a size that the solid angle it subtends at the centre of the test plate is $2\pi/1000$, then the illumination of the test plate will be only $2E/1000$, so that by providing the top of the tube with discs of various known apertures, the brightness of the test plate can be reduced to a corresponding fraction of the illumination it would have received from the whole sky. The method of applying this to the Trotter photometer is shown in *Fig.* 59. Thus only a very small proportion of unobstructed sky is needed, but of course the assumption is made that the sky is uniformly bright all over, and this is usually very far from being the case except with a dull grey sky.

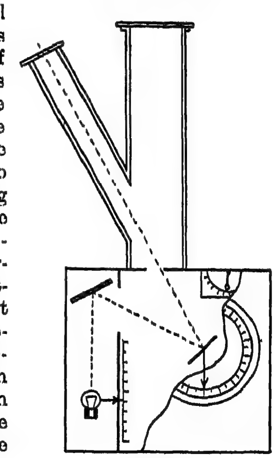


FIG. 59.

It is usual, in making daylight illumination

measurements, to shield direct sunlight from the test card, though the fact that this has been done is generally stated. The average daylight factor for side-lighted rooms of usual construction varies from 0·25 to 1 per cent. A room where the average daylight factor is 4 per cent or over may be considered to possess exceptionally good natural lighting.

VI. Life Testing of Incandescent Lamps

§ (77) Connection between Life and Efficiency.—There are, in the main, two characteristics of an electric glow-lamp which determine the quality of its performance from the economic point of view. The first of these is its efficiency—*i.e.* the candle-power which it gives per watt of electrical energy consumed—while the second is its life, *i.e.* the number of hours for which it may be expected to give this candle-power within certain specified limits. These two characteristics are so intimately connected that if the efficiency be increased by raising the voltage applied to the lamp terminals the life is inevitably reduced. This fact will at once explain both the great importance of the life testing of electric lamps and the necessity for ensuring that such a life test is carried out under the efficiency conditions at which the lamps are to operate in practice.

§ (78) Fall of Candle-power, and "Smashing-point."—In general, it is found that the candle-power of an electric glow-lamp of either the carbon or tungsten filament type falls gradually as the lamp is run (after a short initial period of somewhat uncertain behaviour, generally including a preliminary rise), and this fall of candle-power continues until the filament is fractured either by accidental mechanical shock or by some apparently spontaneous action while in use. It may frequently happen that the fall in efficiency of the lamp, after a certain period of time, reaches the point at which the additional power necessary to give the requisite amount of light is more expensive than the replacement of the lamp by a new one. The point at which this occurs is often termed the "smashing-point" of the lamp, and it is for this reason that it is necessary carefully to define the term "useful life" as applied to an electric lamp. Very frequently this is defined as the time which elapses before the candle-power falls to 80 per cent of its initial value (the voltage being maintained constant throughout the run), or to previous failure, provided this does not take place by accidental means (*i.e.* when the lamp is not burning).

§ (79) Life-test Conditions.—A life test may be made under several different sets of conditions, and these conditions must be carefully specified in order that the desired information may be afforded by the test. The simplest form of test is that in which the lamps are run throughout at constant (generally rated) voltage, all the photometric measurements being made at this voltage. Such a test does not, however, give the most reliable information as to the life performance of a set of lamps. Since life testing must necessarily be by sample, it is essential that the conditions under which the sample lamp is run shall be such as to give the nearest approximation to the average life of the lamps which it represents. This is best attained by a test at definite efficiency, *i.e.* the average working efficiency of the batch of lamps represented by the life-test lamps. This of course involves the adjustment of the voltages on each of these lamps to the values at which they give this definite efficiency. In general, therefore, each life-test lamp must run at a voltage peculiar to itself, and provision must be made for this arrangement in designing any equipment for the life testing of electric lamps. A less satisfactory alternative, when this correct procedure is not possible, is to select as life-test lamps those whose efficiency at rated voltage happens to be nearest to the mean or rated value. These lamps are then tested for life at rated voltage.

§ (80) Forced Life Test.—The life of most modern electric lamps at normal working efficiency is from 500 to 2000 hours, and many attempts have been made to avoid the long delay occasioned by tests such as those described above, and to substitute tests at a higher efficiency. By this means a shorter life is obtained, and then some form of correction factor is applied in order to calculate the life at normal efficiency. Such a test is termed a "forced" life test. The chief difficulty of this method lies in the fact that correction factors differ widely for lamps of different construction, and reliable factors can only be obtained as the result of life tests of large numbers of similar lamps under normal and "forced" conditions. Even with this information the correction factor can only be applied over a comparatively small range of efficiency, but nevertheless a considerable amount of time is saved by adopting this procedure where accuracy is of less importance than speed. The subject of the correction factors applied in forced life tests will be referred to again in the concluding paragraphs of this section.

§ (81) Life-test Installation.—From what has been said above, it will be clear that in any life-test installation two requirements of first importance are (1) a current supply of which the voltage is carefully regulated, and (2) arrangements for applying any desired voltage to each particular lamp on test. The apparatus used for this purpose at the large testing laboratories are generally similar

differing only in details of arrangement. The description here given is of the installation at the National Physical Laboratory,[1] but that of the Bureau of Standards [2] is not greatly different in general principle.

§ (82) VOLTAGE REGULATION.—An alternating current supply of 55 cycles and 240 volts from a dynamo, coupled with a Tirrill regulator, feeds an autotransformer from which leads are taken to a number of racks supported in an iron framework. A diagram of the wiring of one of these racks is shown in *Fig.* 60. The pick-off points of the transformer

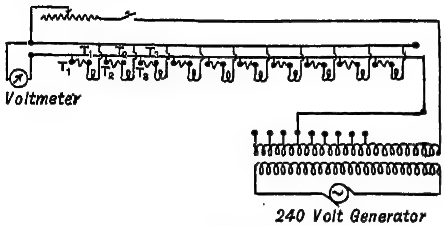


FIG. 60.

permit of any desired voltage in steps of 5 volts being applied to the leads of any rack. This voltage is further adjustable by means of a series resistance so that if lamps are being run at specified voltage they can be put in the sockets on the rack, and the terminals T_1T_1, T_2T_2, etc., can be connected by short pieces of copper wire. More frequently, however, the voltage at the terminals of each lamp on the rack is different, and then small resistances of the form shown at R in *Fig.* 61

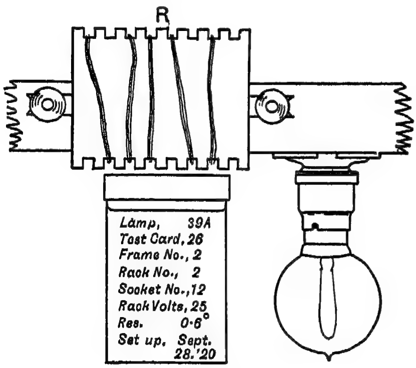


FIG. 61.

are inserted between the terminals T_1T_1, T_2T_2, etc. Each of these resistances is, to the nearest tenth of an ohm, that required to give, with the current taken by the lamp to which it applies, the necessary voltage drop between the rack leads and the lamp terminals.

§ (83) CANDLE-POWER MEASUREMENTS.—Before the lamp is put on life test, measurements of candle-power and current at various voltages are made in the usual way (see § (20)), and the voltage at which the desired efficiency is obtained is then deduced. This voltage then becomes the life-test voltage of the lamp, and no attempt is made to alter this voltage as the efficiency of the lamp falls with lapse of time. Further, the candle-power measurements at stated intervals during the run are made either at this life-test voltage or, more frequently, at rated voltage. These measurements are often made at the expiration of 0, 50, 100, 200, and each subsequent 200 hours after the commencement of the test. In the case of a forced test the intervals at which candle-power measurements are made may be much shorter than this, as the total life is reduced in length.

§ (84) POSITION OF THE LAMPS.—The card shown in *Fig.* 61, close to a lamp, is a very useful auxiliary in practical working. It shows, for the lamp in question, the reference number of the lamp, its position on the life-test rack, the voltage applied across the rack leads, and the resistance necessary to reduce this voltage to the life-test voltage of the lamp. The racks are so arranged on their framework that the lamps can be burnt upright or pendent, the latter being the more usual condition. The racks are inspected at frequent intervals and failures are noted, as far as possible, to the nearest hour. It is the usual practice to regard any lamp, the filament of which fractures when no current is passing through it, as having been accidentally broken. The results on such a lamp are then not included in determining the average life of the group to which it belongs. If the filament of a lamp break, and fall across another portion so as to complete the circuit through the lamp and cause it to burn, that lamp is nevertheless regarded as broken, and removed from the test.

§ (85) LIFE-TEST CURVES.—Lamps are generally run until failure of the filament occurs, or until the candle-power, measured at one of the intervals mentioned above, shows more than 20 per cent drop below the initial value. The interpretation of life-test results is a matter requiring very careful consideration. It is usual to draw the candle-power time curve for each individual lamp, and then to draw two curves showing respectively the average candle-power and the average value of watts per candle for the whole number of lamps burning at any time. Thus in drawing these latter curves, lamps removed from the test, either on account of breakage or candle-power fall, are not included in computing the averages for times subsequent to their removal.

[1] Paterson and Rayner, *Illum. Eng.*, London, 1908, i. 845.
[2] Middlekauff, Mulligan, and Skogland, *Bureau of Standards Bull.*, 1916, xii. 607.

A set of such curves for a batch of six lamps is shown in *Fig.* 62, where the individual candle-power curves are shown on the left, and the mean candle-power and average watts-per-candle curves are shown respectively above and below on the right. The removal of a lamp from the test is indicated by an arrow at the appropriate point on the time scale.

§ (86) Interpretation of Results.—For specification purposes, however, it is desirable to have, in addition to the full information afforded by the curves, some figure of merit for the life-test lamps by which it may be possible to judge of the probable performance of the lamps they represent. The figure adopted by the British Engineering Standards Association in this country is the "test life," defined as the average number of hours burnt by all the lamps in a group throughout a specified running period. It is therefore the total number of hours burnt by all the lamps in a group, throughout each running period, divided by the number of lamps.

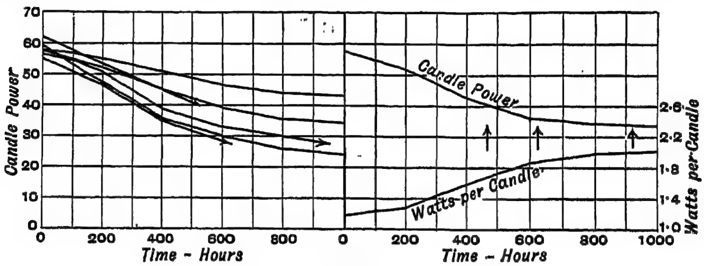


Fig. 62.

§ (87) Forced Life Efficiencies.—The advantages of speed and economy of power possessed by a forced life test were pointed out earlier in this article, as well as the difficulties attending the correct interpretation of such tests. These difficulties are diminished in proportion as the extent of the "forcing" is reduced. The extensive life tests carried out at the Bureau of Standards[1] are made at efficiencies corresponding to about 0·9 to 0·95 watts per candle for tungsten lamps rated at from 1 to 1·15 watts per candle, and in many other cases the use of forced life tests is customary.

The determination of the conditions at which a forced test is to be run may be made in two different ways. The first method is to measure the voltage at some stated watts per candle, and to multiply this by a constant factor to determine the forced test voltage. A forced test voltage in common use with this method in America is 130 per cent of the voltage at which the lamp absorbs 1·23 watts per candle.[2]

The second method is to run the lamps at the voltage which gives them, initially, some specified watts per candle higher in efficiency than the normal by a constant amount.

The life-test voltage having been determined as above, it is usual to make the photometric measurements at rated voltage for the sake of consistency and simplicity in records. Since the life test is considerably shortened in length of time taken, the intervals between the photometric measurements are correspondingly reduced.

§ (88) Reduction of Forced Life Results.—It is usual to reduce the life results of a forced life test to those of a test at normal efficiency by means of a relationship such as the following:

$$(\text{Life}) = (\text{constant}) \times (\text{watts per candle})^n.$$

For although Cady[3] and Edwards[4] have shown that n is not constant over a wide range of variation of the efficiency, yet within the limits of 15 per cent in voltage above and below normal the error in the computed life will not be serious if the departure of the efficiency from the normal value is not too great. Lewinson has shown[5] how much the value of the exponent n depends on the construction and type of the lamp under test. For 100-watt lamps of various constructions this exponent varied from 6·2 to 8·2 with a mean of 7·4. For lamps of various sizes the value of n was also found to vary according to the following table:

Watts.	n.	Watts.	n.
15	6·8	40	7·4
20	7·2	60	7·5
25	7·3	100	7·8

Middlekauff, Mulligan, and Skogland[6] use

[1] Middlekauff, Mulligan, and Skogland, *loc. cit.*; and *Am. Illum. Eng. Soc. Trans.*, 1915, x. 814.

[2] Lewinson, *Am. Illum. Eng. Soc. Trans.*, 1916, xi. 817.

[3] *Am. Illum. Eng. Soc. Trans.*, 1908, iii. 459.

[4] *Gen. Electric Rev.*, 1914, xvii. 282.

[5] *Loc. cit.* p. 822.

[6] *Loc. cit.* p. 826.

the value 7·4 for all sizes of tungsten lamps from 25 watts upwards, and the value 5·83 for carbon lamps.

The same value of n cannot be applied to all interpretations of a forced life test. The figures given above refer to a "useful life" interpretation, *i.e.* life to 80 per cent of initial candle-power or earlier burn-out. For the 60-watt lamp the value of n given by Lewinson for a "total" life interpretation—*i.e.* life to burn-out—is 6·6, while for an 80-per-cent life—*i.e.* life to 80 per cent of initial candle-power with all burn-outs eliminated—it is 7·6 instead of the figure of 7·5 given in the table above.

An important factor in the loss of candle-power during the life of a lamp is the blackening of the bulb, especially if this be small compared with the size of the filament. Some lamps contain bulb-blackening preventives, which depend for their effective operation on the temperature of the bulb. It will be clear that if the variation of the action of this preventive with change of temperature is different in two classes of lamps the life efficiency factor will be different also, and this effect requires consideration when forced life tests are being made.

§ (89) Special Life-test Conditions.—What has been said above applies, in the main, to life tests of all classes of lamps, but in the case of lamps required for use under special conditions it is often desirable that the life test should be carried out as nearly as possible under the conditions of use of the lamps represented. Thus, for example, lamps to be used in confined spaces should be represented on life tests by lamps burning in similar-sized spaces. Lamps to be run off small-capacity portable accumulators, where the lamp may be subjected to appreciable over-voltage at the beginning of each run, may be required to be tested for life on a circuit in which this over-voltage is imitated. Special life tests of this nature require special arrangements of the electrical circuits.

§ (90) Number of Lamps required for Life Test.—It is also desirable that the number of life-test lamps should be a larger percentage of the total number of lamps represented than is the case with more normal tests. The standard specification of the British Engineering Standards Association calls for a life test on at least one-half of one per cent of the lamps in a batch, with a minimum of five lamps, in the case of ordinary tests, and this number should be doubled or trebled in the case of special or of forced tests.

§ (91) Dependence of Life on Voltage.—When setting up lamps on life test it is, of course, necessary to ensure that failure of one lamp does not entail excess voltage on any of the others. Thus life-test lamps must not be set up in parallel on a circuit containing any appreciable external series resistance, otherwise the failure of one lamp will increase the effective resistance of the lamp portion of the circuit, with consequent rise of voltage, to act adversely on the life performance of the remaining lamps. Constancy of voltage is, in fact, one of the chief requirements of a reliable life test, and the B.E.S.A. Specification, referred to above, calls for a limit of variation which shall not exceed 1 per cent as regards momentary fluctuations, or what is appreciable on an ordinary large-scale type indicating voltmeter as regards permanent error in the voltage at which the lamps are run. The reason for this requirement is readily understood from the high value of the life efficiency characteristic when it is remembered that this means a life-voltage exponent of between 15 and 20 for a vacuum-type tungsten-filament lamp. The degree of dependence of life upon efficiency for gas-filled lamps has not yet been determined, and it will probably prove to be extremely variable owing to the large number of independently variable conditions in this type of lamp.

§ (92) Effect of Alternating or Continuous Current.—It has been customary to consider that a life test may be run on either continuous or alternating current, and experiments by Mongini [1] show that the results obtained by both methods are practically identical. The same conclusion is reached by Merrill, Cooper, and Blake,[2] who found that actual switching on and off did not appreciably affect the life of the lamps. If, therefore, it be assumed that the only effect which might have a tendency to decrease the life on alternating current is that of the rush of current which takes place when a metal filament lamp is first switched on (due to the lower resistance of the filament when cold), the smallness of the effect on the life found for actual switching on and off can only be diminished when the current impulses follow one another so rapidly that the degree of cooling in the filament is comparatively slight.

VII. Heterochromatic Photometry

§ (93) Effect of Colour Difference in Photometry.—In the other sections on photometry it has been assumed throughout that the two surfaces whose brightnesses are being compared appear to the observer to be of the same colour. Such a condition is, however, the exception rather than the rule in practical photometry, except in cases where the work is confined to the measurement of sources which are very uniform in character. The eye is a very sensitive judge of colour difference, and when two exactly similar

[1] *Atti dell' Assoc. Elettr. Ital.*, 1913, xvii. 690.
[2] *Am. I.E.E. Proc.*, 1910, xxix. 945.

surfaces are seen side by side in a photometer (as, for example, in the Lummer-Brodhun head), if they are illuminated by two tungsten lamps operating at efficiencies differing by as little as 2 per cent, a colour difference is just perceptible to the eye of a practised observer.

Actually, in such a case, the light given by the less efficient lamp is very slightly yellower than that given by the other, but the appearance in the photometer head is, by contrast, that of a slightly pink patch on a bluish field and *vice versa.* Even such a very slight colour difference as this is sufficient to reduce quite seriously the accuracy of a photometric balance. The colour difference makes it impossible to obtain exact equality between the two halves of the field, and the eye has to allow for the difference in *hue* when endeavouring to obtain a balance of *brightness.* Of course much more serious differences are met with in practice when comparing electric glow-lamps with flame standards, or when measuring gas or acetylene sources by means of electric glow-lamp sub-standards. When it is a case of measuring daylight illumination, or the light from a high-intensity electric arc by means of tungsten-filament sub-standards, direct comparison becomes very inaccurate indeed, and the majority of observers will, on different occasions, obtain readings differing by as much as 20 per cent. Further, different observers do not obtain results in agreement with one another, and it is therefore necessary that special methods shall be adopted for the comparison of lights in which the colour difference exceeds even a small amount.

§ (94) Methods of Heterochromatic Photometry. — Theoretically no physical equality can ever be obtained between lights of different colours, because the things being compared differ in kind as well as in degree. But physiologically it is a matter of experience that, provided the difference in kind be not too great, equivalence in degree can be established within assignable limits by observers having normal vision For even when signal-green and ruby-red lights are being compared, it is possible to raise the brightness of the green to such a degree that no doubt is left in the observer's mind that the green is definitely the brighter of the two, while similarly there is a much lower intensity at which the red can quite confidently be asserted to be the brighter. The aim of heterochromatic photometry is to reduce these limits as much as possible for the cases met with in practical photometry, and this problem has been attacked, in the main, along three lines, viz.: (i.) the flicker method, (ii.) the use of coloured glass or gelatine filters or solutions, and (iii.) the division of the colour difference to be dealt with into a number of smaller colour steps.

§ (95) The Flicker Photometer. — The flicker method depends upon the phenomenon of visual diffusivity in the human eye (see article on "The Eye," § (19)). When two bright surfaces are presented to the eye in rapid alternation a flicker is perceived, the degree of which depends both on the rapidity of the alternation and also on the identity of the two surfaces as regards brightness and colour. The more nearly identical the surfaces the slower the speed at which flicker ceases to be perceptible, and the principle of the flicker photometer lies in producing a rapid alternate presentation of the two comparison surfaces to the eye of the observer, and the adjustment of their relative brightnesses until no flicker is observed at a comparatively low frequency of alternation.

§ (96) The Whitman Photometer. — Several instruments have been designed on this principle. That of Whitman [1] is shown in plan in *Fig.* 63. C is one comparison

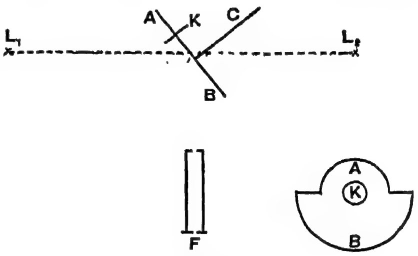


Fig. 63.

surface, and AB the other. AB consists of a disc of the shape shown in the right-hand bottom corner of the figure. Rotation of AB about the axis K causes the eye at F to see first one and then the other comparison surface in succession. A photometric balance can be repeated with this instrument to about 2 per cent when using lights of the same hue. The accuracy is of course less with different coloured lights.

§ (97) The Bechstein Photometer. — Rood employed a plano-concave cylindrical lens oscillating in front of a Ritchie wedge so that first one side and then the other was brought into the field of view. This principle has been developed by Bechstein,[2] who uses the lens and prism system shown at L in *Fig.* 64. In the position shown in the figure the eye at A sees a circular field of which the outer annular portion is due to the right-hand side of the prism P, while the inner circular portion is due to the left-hand side of P. If now L be rotated as a whole through 180° the

[1] *Phys. Rev.*, 1895–96, iii. 241.
[2] *Zeits. Instrumentenk.*, 1916, xxvi. 249

outer and inner parts of the field are respectively illuminated by the left- and right-hand sides of P, so that as L is rotated by a small

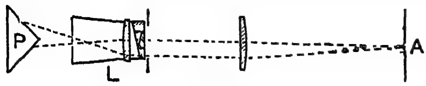


FIG. 64.

electric motor an alternation of contrast is obtained in the field of the photometer.

§ (98) THE SIMMANCE-ABADY PHOTOMETER.—In the flicker photometer of Simmance and Abady [1] a plaster disc consisting of a combination of two truncated cones is used. Its formation may be best understood from *Fig.* 65. ABCD and EFGH are two exactly similar truncated cones, divided respectively by the planes AC and EG. The portions ABC and EGH are removed, and EFG is then placed on ACD so that the resulting solid has the form shown in *Fig.* 66, which represents it as seen edge-on in four positions 90° apart. It will be clear that if the two sides of such a disc be illuminated by the two sources to be compared, the line of demarcation will swing back and forth across the field of view for every rotation of the disc, and thus a

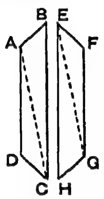


FIG. 65.

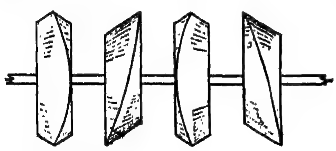


FIG. 66.

flickering field will be obtained as before. Discs designed on an exactly similar principle, and giving 4 or 8 alternations instead of 2 for each rotation, have been designed by Krüss.[2]

§ (99) THE WILD PHOTOMETER.—The flicker photometer designed by Wild [3] consists of a Bunsen disc in which a semicircle or two quadrants are waxed, the remainder of the disc being plain. It is mounted so as to be perpendicular to the direction of the beams of light to be compared, and both sides of the disc are viewed simultaneously by means of mirrors. Rotation of the disc by clockwork, or by a small electric motor, produces the field alternation. The criterion in this photometer is not absence of flicker, but equality of flicker on both sides of the field. It therefore possesses the advantage that the appearance of the field, when out of balance, indicates the direction in which the head has to be moved. This instrument has been stated to have a sensitiveness of 0·5 per cent with lights of the same colour and 0·9 per cent when comparing red and green lights.

§ (100) EFFECT OF SPEED ON SENSITIVITY OF FLICKER.—The speed of a flicker photometer has a very noticeable influence on its sensitivity. Dow [4] has investigated this problem and finds that for lights of different colours the speed giving maximum sensitivity is much higher than it is for lights of the same colour. The range of speed over which maximum sensitivity is obtained is also more restricted in the case of different coloured lights. The most favourable speed varies both with the illumination and with the difference of colour of the two fields, the results obtained when comparing green and white lights being shown graphically in *Fig.* 67. The abscissae are frequencies of field

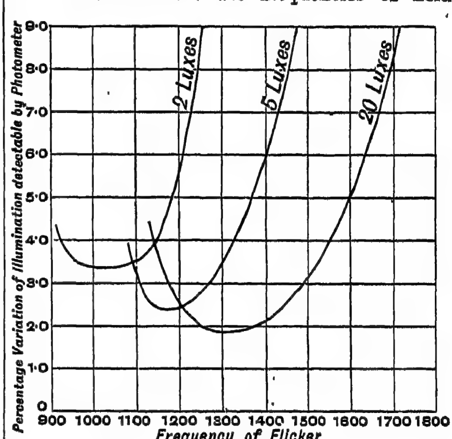


FIG. 67.

alternation (*i.e.* the number of changes per minute), while the ordinates show the percentage change of illumination which can be made without flicker, *i.e.* the difference between the illumination ratios at the two positions at which flicker just begins to appear. The accuracy of setting can be made much closer than this, since the mean of the two positions at which flicker is just perceptible may be taken as the position of balance. For illuminations greater than 20 metre-candles the range of sensitivity is approximately the same as for that illumination, while it can also be assumed that a colour difference less than that of the experiments will give a greater range of speed for maximum sensitivity, and that the actual speeds will be lower, tending to limits of 500

[1] *Phys. Soc. Proc.*, 1903, xix. 39.
[2] *Zeits. Instrumentenk.*, 1905, xxv. 98.
[3] *Illum. Eng.*, London, 1908, i. 825.
[4] *Electrician*, 1907, lix. 255.

to 100 for lights of the same colour when the illumination is 2 metre-candles.

§ (101) THE COLOUR-FILTER METHOD.—A second method by which it has been proposed that lights of different colours should be compared is that involving the use of some coloured medium which will either (*a*) bring the hue of one light to approximate equality with that of the other, or (*b*) enable a comparison to be made of the relative intensities of both lights at some particular part of the spectrum. To the former class belong the "photometric" gelatine filters devised by C. K. Mees.[1] In § (6) "Visibility" of the article "The Eye" (*q.v.*) it has been stated that a curve may be constructed showing the relative brightnesses which the average eye will assign to equal quantities of energy at different parts of the visible spectrum. This curve, called the visibility curve, is shown in *Fig.* 2 of that article, and if its ordinates be multiplied by

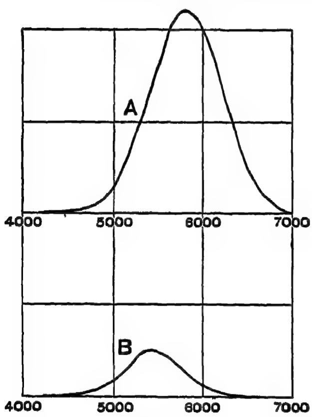


FIG. 68.

the relative energy emissions in the different wave-lengths, of a given source of light, a curve will be obtained showing the relative contributions of the different parts of the spectrum given by this source to its total luminosity. Such a curve for a black body at a temperature of 2000° K is shown in *Fig.* 68, curve A.

Now by means of a spectrophotometer (see article on "Spectrophotometry") the transmission ratio of a coloured transparent medium may be measured for light of any wave-length. If, then, the ordinates of curve A in *Fig.* 68 be multiplied by the corresponding transmission ratios of a colour filter, a curve will be obtained showing the luminosity reaching the eye through the filter. Such a curve is shown in *Fig.* 68, curve B. The ratio of the areas of the two curves A and B gives the over-all transmission ratio of the filter for light of the spectral distribution assumed, and hence this value of transmission ratio can be assumed for all sources the spectral distribution of the light from which approximates to that of the curve used for the calculations.

§ (102) COMPARISON OF DIFFERENT COLOURED LIGHTS.—In this way a tungsten-filament lamp, or even daylight, may be compared with a flame standard by using a blue colour filter, which, when placed between the standard and the photometer, will produce a colour match in the photometer head. A knowledge of the energy distribution curve of the standard and the visibility curve of the average eye gives at once the luminosity curve of the standard. This and the transmission ratio curve of the filter give the over-all transmission ratio of the filter for the light from the standard, and hence the candle-power of the combination of standard and filter. Instead of a blue filter in front of the standard it is often desirable to use a yellow filter in front of the light of higher efficiency. In this way an undue reduction of candle-power on one side of the photometer may be avoided.

The above process of finding the transmission ratio of a colour filter is very laborious, and a less fundamental but much simpler method is that employed at the Bureau of Standards and at the Physikalisch-Technische Reichsanstalt. In this method it is assumed that the mean value of candle-power obtained by a large number of observers with a direct comparison photometer, such as the Lummer-Brodhun, approximates very closely to the true value even when the colour difference involved in the comparison is considerable. The transmission ratio of a colour filter is determined by comparing the candle-power of a given lamp (of Hefner flame colour) with a standard (*a*) without the filter and (*b*) with the filter placed between the lamp and the photometer head. The ratio of the candle-power in case (*b*) to that in case (*a*) is then assumed to be the transmission ratio of the filter for light of that colour, and the combination is used for the determination of candle-power of test lamps as usual.

The advantage of this method over a direct comparison involving colour difference in every case is that the determination of transmission ratio can be made by a large number of observers, and when this has been done, the photometry of test lamps involves no further colour difference, so that a much smaller number of observers is sufficient.

§ (103) THE CROVA WAVE-LENGTH METHOD.—The second alternative, of using a colour filter having only a narrow band of transmission in the yellow part of the spectrum, was suggested by Crova in 1881.[2] He found

[1] *Amer. Illum. Eng. Soc. Trans.*, 1914, ix. 990.

[2] *Comptes Rendus*, 1881, xciii. 512.

that if the visibility curves of various sources having approximately the same spectral distribution as a black body at various temperatures were plotted with equal total areas, these curves all intersected, *i.e.* had equal ordinates, at the wave-length 0·582 μ. It therefore followed that a comparison of such lights at this wave-length gave a correct comparison of their total candle-powers. This wave-length, termed the Crova wave-length, has been redetermined and probably lies more nearly in the region 0·557 μ to 0·562 μ. One great advantage of this method lies in the fact that any filter which has only a narrow band of transmission must necessarily reduce the intensity of the light very greatly. Crova suggested the use of a combination of solutions of perchloride of iron and chloride of nickel in a glass vessel 7 mm. thick. This combination transmits between the wave-lengths 0·630 μ and 0·534 μ with a maximum at 0·582 μ.

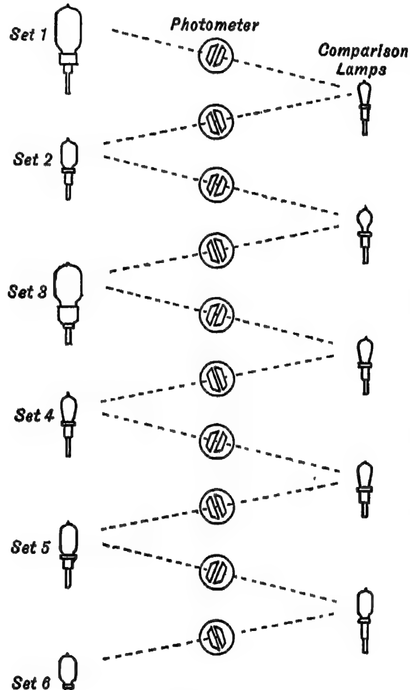


FIG. 69.

§ (104) THE CASCADE METHOD. — The third method of colour photometry does not attempt to eliminate the colour difference, but simply divides it into a number of small steps. This is the method employed for the measurement of the electric sub-standards used at the National Physical Laboratory. Between the flame standard, in this case the Vernon-Harcourt pentane lamp, and the highest efficiency sub-standards used, viz. those operating at 0·67 candles per watt, four sets of tungsten or carbon filament sub-standards are interposed, as shown diagrammatically in *Fig.* 69. The lowest efficiency set consists of tungsten-filament lamps operating at 0·11 mean horizontal candles per watt and the light from these matches that from a pentane lamp; the other lamps are respectively (1) single carbon loop-filament lamps operating at 0·20 m.h.c. per watt; (2) double horse-shoe carbon filament lamps operating at 0·26 m.h.c. per watt; (3) metallised carbon filament lamps operating at 0·33 m.h.c. per watt; (4) tungsten grid filament lamps operating at 0·52 m.h.c. per watt, and (5) tungsten grid filament lamps operating at 0·67 m.h.c. per watt.

These efficiencies are such that the colour difference between any two neighbouring sets is approximately the same throughout the series, and each set of lamps is compared with the set below it by not less than six observers, each taking at least 30 observations on each lamp. The observers work in pairs, every one of the possible combinations of observers being employed. Thus the values obtained by each observer on each lamp are taken on five different days. By this method the effect of colour difference on the comparison is minimised, and a direct comparison of set 6 with set 1, made in an exactly similar way to that described above, shows that, in fact, the sum of the probable errors of the different steps in the comparison by what is called the "cascade" method is slightly less than the probable error of the direct comparison.

There is, however, another much greater practical advantage in the cascade method. Ordinary photometric comparison is made by not more than two observers, so that in the case of a considerable colour difference the chance of any two observers obtaining a result in agreement with that obtained by a much larger number has to be considered. The intercomparison of the sub-standards, on the other hand, is carried out by at least six observers, and then for any subsequent photometry a sub-standard is available of a colour quite close to that of the test lamp. The small remaining colour difference is then all that remains to be considered when assigning the accuracy of comparison by two observers.

§ (105) THE COMPARISON LAMP IN CASCADE. —It is often the practice, when working with small colour differences, to use a comparison lamp giving light of a hue midway between that of the test lamp and that of the nearest available sub-standard. In this way the colour difference is halved, but as the same difference appears on opposite sides of the photometer in standardisation and in the test-lamp

measurement, it is doubtful whether any increase in accuracy results from this procedure. When working with sources far bluer than the highest efficiency sub-standard a "half-way" colour filter may be used on the comparison lamp side, or the method adopted by the Bureau of Standards (§ (102)) must be used.

§ (106) Other Methods of Heterochromatic Photometry.—Other methods, not depending on any of the three principles above described, have been proposed and may be briefly noticed. Macé de Lépinay suggested placing in front of eyepiece of the photometer first a green and then a red solution. The ratio of the intensities found in these two cases gave, by means of a table, the intensity for the whole spectrum. Wybauw used a "compensation" photometer in which the light from the (more powerful) source to be measured illuminated one side of the photometer, while the other side received light from the standard and, in addition, a fraction of the light from the source to be measured. Thus the hue difference was reduced, but the sensitiveness of the measurement was also diminished.

§ (107) The Visual Acuity Method.—Another method, which has been proposed by Weber and others, depends on the amount of light necessary for the visual perception of detail. Patterns consisting of concentric black circles of different thicknesses, or a series of fine dots of progressively diminishing size, are illuminated in turn by means of the lights to be compared, and it is assumed that equal ease of discrimination corresponds with equality of illumination for the different lights. The theoretical soundness of this form of colour photometer has been frequently called in question, and it appears more than doubtful whether visual acuity can be used in this way as a judge of equality of brightness without reference to the colour of the light used.

§ (108) The Contrast Pattern Method.—A colour photometer devised by Von Czudnochowski [1] depends upon the production of two sets of shadows formed by mutually perpendicular wire gratings, each set being illuminated by one of the sources to be compared. The field produced consists of two shadow patterns which appear grey on a white background, with dead black intersections. If one of the sources is moved sensibly out of balance one of the patterns becomes coloured.

§ (109) Effect of various Factors on Flicker Sensitivity.—The various phenomena underlying the comparison of lights of different colours have been the subject of investigation by many workers. Notably J. S. Dow [2] has shown that, while a single observer may obtain very consistent results on a single occasion when using an equality of brightness photometer to compare lights of different colours, yet the same observer, after a few hours' rest, may obtain another series of readings, equally consistent among themselves, but differing from the first set by 5 or even 10 per cent. Dow also finds that the readings are affected by the size of the image produced on the retina by the photometric field. It results from this that the size of the surfaces used for the purpose of comparison, as well as the distance of the eye from these surfaces, are of importance in heterochromatic photometry. He found in a particular case of comparison between red and green lights that there was a 25 per cent difference in setting according as the telescope of a Lummer-Brodhun photometer head was in its extreme "in" or "out" position. He also investigated the effect of the Purkinje phenomenon in colour photometry, and found that this had no appreciable effect on illuminations above 1 metre-candle, so that it may be neglected in ordinary photometry. The experiments, however, seemed to show that the effect was less with the flicker than with the equality of brightness photometer.[3]

§ (110) Comparison of Flicker and Equality of Brightness Methods.—Ives carefully investigated the relative merits of the different methods of colour photometry,[4] and came to the conclusion that the flicker method was more sensitive than the equality of brightness method, and that the results obtained by its means were more reproducible in the case of lights differing considerably in hue. This opinion was confirmed by Crittenden and Richtmyer,[5] but they found that for sources having a relatively high intensity at the blue end of the spectrum, the values given by the flicker photometer departed appreciably from those given by the contrast type, the difference being estimated at as much as 3 per cent for such sources as the gas-filled lamp. They also came to the conclusion that for individual observers, or for small groups, the flicker photometer gave a result which was closer to the mean than that found with the contrast instrument. Reference [6] should also be made to the work of Middlekauff and Skogland.

A critical discussion of the relative merits of the two methods has been given by Hyde,[7] who, in his determination of the visibility curve of the human eye, preferred the equality of contrast method as being the most suitable.

§ (111) Colour Difference in Illumination Measurements.—The above description has been practically confined to ordinary

[1] *Illum. Eng.*, London, 1908, i. 283.
[2] *Phil. Mag.*, 1906, xii. 120.
[3] *Phil. Mag.*, 1910, xix. 66.
[4] *Ibid.*, 1912, xxiv. 149.
[5] *Amer. Illum. Eng. Soc. Trans.*, 1916, xi. 353.
[6] *Bureau of Standards Bull.*, 1918, xiii. 287.
[7] *Astrophys. J.*, 1918, xlviii. 65.

photometry, but in the measurement of illumination the problem of colour difference is even more important on account of the many different types of illuminants met with, often in the same building. Further, the brightness of the test surface may sometimes fall below the limit at which the Purkinje effect cannot be neglected, so that colour difference between the light to be measured and the comparison source may cause considerable errors at these lower values of illumination. Above all, the colour differences met with in ordinary photometry never approach the difference experienced when measuring daylight illumination by means of a portable photometer in which the comparison lamp is an ordinary tungsten-filament vacuum glow-lamp. The use of a colour filter is almost universal for daylight illumination measurements. This may take the form of a yellow filter placed between the test surface and the photometer, or a blue filter used in front of the comparison lamp. The latter scheme has the disadvantage that it reduces the upper limit of working of the photometer. Instead of interposing a yellow filter between the test surface and the photometer, a yellow test surface is sometimes used. The constant of such a surface, or filter, can be found on the photometer bench by a number of observers, and it may then be used in the portable instrument by a single observer. The transmission ratio of a yellow filter must, of course, be determined for light of daylight colour. It should be noticed, in this connection, that the colour of daylight is by no means constant. The colour of the light from a clear blue north sky is much richer in blue rays than the light derived directly from the sun or from white clouds illuminated by it. The difference, however, is not sufficient to invalidate the transmission ratio of a yellow filter determined for light of the colour of sunlight.

§ (112) Determination of Filament Temperature.—The ease with which the eye can distinguish colour difference in an ordinary equality of brightness photometer has been used by C. C. Paterson and B. P. Dudding [1] for the determination of the true temperature of an incandescent body. They found that the light from an ordinary tungsten or carbon filament vacuum lamp approximated very closely in spectral distribution to that of a "black body," and that a close and invariable relationship existed between its efficiency, expressed in candles per watt, and its temperature. Thus if one side of a photometer be illuminated by such a lamp while the other side receives light from an incandescent body whose radiation has the same spectral distribution as a "black body," the efficiency of the lamp may be altered, by changing the voltage applied to it, until the two sides of the photometer appear to be of exactly the same colour. An exact equality of brightness must be maintained by moving the photometer between the sources during the process of obtaining this colour match, and the temperature of the incandescent body may then be assumed to be very closely the same as that of the lamp filament. The lamp filament may be calibrated by an exactly similar method against a standard "black body," but the authors above quoted (*loc. cit.*) have found that for all drawn tungsten-filament lamps of ordinary vacuum type and disposition of filament, the relation between temperature and efficiency follows the relationship

$$\log\frac{L}{W}=C-m\log_{10}T-185\log_{10}\left(1+\frac{115}{T}\right),$$

where L/W expresses the efficiency of the lamp in lumens per watt, T is the absolute temperature, and C and m are constants having the values 21·51, 4·58 and 23·31, 5·1 respectively for carbon and tungsten filaments. They have also found that the watts consumed by a lamp vary with the absolute temperature according to the relation $\log_{10}W=C_1+4{\cdot}58\log_{10}T$ and $\log_{10}W=C_2+5{\cdot}1\log_{10}T$ for carbon and tungsten filaments respectively.

VIII. The Photometry of Projectors

The photometry of light-projection apparatus falls into a class by itself on account of the many difficulties involved and the special means which have to be employed in order to overcome them. At the same time it is of the utmost importance to obtain information as to the relative performance of different types or patterns of apparatus such, for instance, as searchlights, motor-car headlights, signal lights for marine or land service, and similar special optical devices.

§ (113) Difficulties due to Beam Concentration.—The chief difficulties met with in these tests arise from the fact that the light does not diverge from a source of which the dimensions may be neglected in comparison with the distance from it at which the measurements are made. In all the cases mentioned above, the light from the source is redistributed by optical devices, and it is therefore necessary to ensure that the measurements are made at such a distance from the apparatus that the inverse square law may be assumed to hold within the accuracy desired. It is not necessary, of course, that distances should be measured from the source itself, and often it is assumed that the optical centre of the device lies at the meeting point of the

[1] "The Estimation of High Temperatures by the Method of Colour Identity," *Phys. Soc. Proc.*, 1915, xxvii. 230.

extreme rays of the projected beam. This assumption, however, is generally no more than a convenient approximation to the truth, for it cannot always be assumed that the light is emitted in all directions from a single point. Often the light emitted in two different directions may behave as if it emanated from points which are separated by a distance far from negligible in comparison with the distance at which measurements have to be made.

§ (114) Minimum Distance for Photometric Measurements.—It may be generally assumed that the inverse square law holds for distances greater than fifty to four hundred times the diameter of the optical aperture with beams of 20° to 2° total divergence. As a very approximate guide it may be assumed that the inner limit of distance at which the beam has attained its final distribution is given by Kd/θ, where θ is the total angle, measured in degrees, of the cone of light formed by the beam, d is the diameter of the aperture, and K is a constant lying between 600 and 1000.

This rule leads to the result that for such apparatus as motor-car headlights, where the diameter of the mirror is of the order of 10 inches and the divergence may be as little as 5°, photometric measurements should always be made at least 100 feet away from the headlight. On the other hand, for a lens such as that used in a ship's navigation light, where the divergence may be as much as 20° with a lens height of 7 inches, a distance of about 20 feet is sufficient. In both these cases the chief difficulty is that of obtaining sufficient light to enable measurements to be made at these comparatively great distances, and very often a compromise has to be effected by making measurements at two shorter distances, and obtaining an approximation to the desired result by an extrapolation.

§ (115) "Effective" Candle-power.—The information usually desired is that given by a curve of distribution of illumination on a screen placed so as to be perpendicular to the axis of the beam at a convenient distance from the source. This distribution may often be conveniently found by actual measurement of brightness at different portions of a white screen placed in the path of the light, using a form of portable illumination photometer (see p. 441). It is often more convenient, however, to keep the photometric apparatus fixed in position and to move the source either in altitude or azimuth. Measurements of illumination can then be made by means of a photometer head fixed in a given position, with a comparison lamp movable along a bench directed away from the source. Alternatively, the test surface of a portable photometer (see p. 441) may be fixed in a convenient position, and measurements of illumination at this position may then be made for any desired orientation of the projector. The results may be expressed either directly in terms of illumination, or, by calculation, in terms of the candle-power which would be required of a point source placed in the position of the projector in order that it might produce at the screen the illumination actually measured there. The latter figure is generally termed the "effective candle-power" of the source in the direction considered.

§ (116) Distribution Curve for Beams.—Whichever method of expressing the results is employed, the distance from the source at which the measurements have been made should always be stated. The results may be exhibited graphically by means of a curve in which the abscissae represent either illumination at a given distance or effective candle-power, while the ordinates are the corresponding angles of deviation from the axis of the apparatus. Such a curve, for a motor headlight beam of small divergence, is shown in the upper diagram of *Fig.* 70. The lower

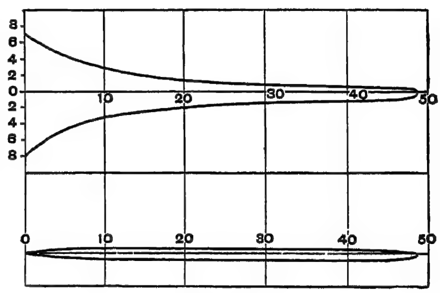


Fig. 70.

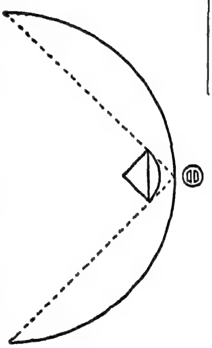
diagram is a polar curve for the same beam, and illustrates strikingly the failure of a polar diagram to give an intelligible representation of light distribution from any form of projection apparatus.

For photometry of small projection apparatus, such as motor headlights, ships' light lenses and hand signalling lamps, the distances required by the formula given above do not generally exceed 100 feet, and consequently measurements can be carried on in the laboratory, where all that is required in addition to ordinary photometric equipment is a tilting table for movement in altitude, and a horizontal turn-table for variation of angle of azimuth. If the beam is roughly symmetrical about its centre, sufficient information is generally given by a curve, such as that shown in *Fig.* 70, representing the mean of measurements made across the horizontal and vertical diameters of the beam. If the beam is not symmetrical in shape, similar curves are obtained along other specified lines of

traverse. Occasionally the patch of screen illuminated by the beam is divided into squares, and the illumination on each of these squares is measured and noted on a figure representing the appearance of the patch.

§ (117) Atmospheric Absorption.—In the photometry of searchlight projectors matters are very different. The divergence of the beam may be as little as 2 to 3°, while the diameter of the mirror is from 2 to 4 feet. Consequently, distances of at least 500 to 1000 feet are necessary for accurate beam tests, and in practice it is customary to employ distances of one-half to two miles. At such distances as these, attainable only in the open, atmospheric absorption cannot be neglected. Even a slight ground mist may cause errors of as much as fifty per cent, which are by no means constant from hour to hour, or even from minute to minute.

The effect of atmospheric absorption may be allowed for in one of three ways. First, a "standard" searchlight beam of known constant characteristics may be used. This beam directed towards the distant measuring station at intervals throughout a test will give, by measurement of its candle-power, the correction to be applied on any given night to the observations made on the other searchlight beams tested during that night. Such a beam may be that given by a large size tungsten arc, or a steady carbon arc burning under standard conditions. In either case the source of light must be used in conjunction with a given parabolic reflector as no two reflectors can be relied upon to give exactly the same distribution of light in the beam.



§ (118) The Telephotometer.—A second method depends on the use of a telephotometer, in which a simple double convex lens forms an image of a large screen (situated at the observing station) on the centre of a Lummer-Brodhun cube. This cube forms part of a photometer of ordinary construction at the station where the searchlight is placed. Simultaneous readings of the brightness of the screen as measured by the telephotometer and by an ordinary portable photometer at the observing station give at once the atmospheric absorption when the calibration of the telephotometer is known.

§ (119) The Two-station Method.—The third method is more direct than either of the foregoing. In this, two observing stations are used at known distances d_1 and d_2 from the searchlight. The light is directed first to one station and then to the other, and measurements of the illuminations are made. If these be I_1 and I_2, and t the transmission coefficient of the atmosphere per unit length (assumed to be the same throughout the region over which the measurements are made), while C is the effective candle-power of the searchlight, then

$$I_1 = \frac{C}{d_1^2} t^{d_1} \quad \text{and } I_2 = \frac{C}{d_2^2} t^{d_2},$$

so that

$$\frac{I_1 d_1^2}{I_2 d_2^2} = t^{(d_1 - d_2)}.$$

If for convenience $d_1 = 2d_2$, then $C = I_2^2 d_2^2 / 4I_1$.

§ (120) The Integrating Hemisphere.—Atmospheric absorption is only one out of many difficulties attending the measurement of candle-power distribution in the beam of a searchlight,[1] and for many purposes sufficient information is obtained by a measurement of the total luminous radiation in the beam. For this purpose some form of integrating photometer must be employed. Either the arc may be placed out of focus, so as to cause the beam to converge on to a white reflecting surface placed inside an integrator of the sphere or cube type, or the beam may be directed on to a hemispherical integrator as illustrated on the left-hand side of *Fig.* 71. This

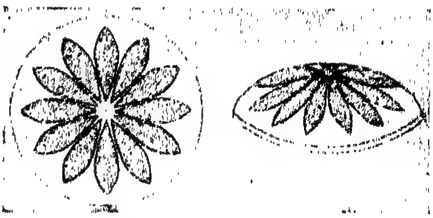


Fig. 71.

integrator consists of a hemispherical matt white surface, furnished at its centre with an aperture and photometer, and having in front of the aperture a convex mirror with foliated black screens as shown enlarged on the right of the figure. These screens are so designed as to produce on the photometer window an illumination proportional to the total flux received by the hemisphere. A full description of the method has been given by Benford.[2]

§ (121) The Focometer.—Another difficulty in the beam testing of searchlights is the necessity for ensuring that the arc crater is kept constantly in the same position with respect to the mirror. This can be done either with a special focometer fitted to the side of

[1] Paterson, Walsh, Taylor and Barnett, *Inst. Elect. Eng. Journ.*, 1920, lviii. 83. F. A. Benford, *Gen. El. Rev.*, 1919, xxii. 668.

[2] *Am. Illum. Eng. Soc. Trans.*, 1920, xv. 19.

the projector case, or by observation of the divergence of the resulting beam. The measurement of the surface brightness of the positive crater of a carbon arc is of very great importance, since, for a given angle of divergence, this quantity determines the brightness of the illumination produced by the beam at a given position. A method by which this quantity may be measured has been indicated in Part II. on "Photometric Standards" (§ 11). All searchlight photometry is affected by difficulties of colour difference, and the use of coloured glasses is general on this account (see § (102)).

IX. The Photometry of Fluctuating Sources of High Candle-power

§ (122) Photometer for Instantaneous Reading.—It is sometimes desirable to obtain an approximate estimate of the candle-power of intermittent or rapidly fluctuating sources such as flares, parachute lights, landing lights, etc. The photometry of such sources presents the special difficulty that photometric balance cannot be obtained except for a brief instant of time. Consequently any photometer involving the use of moving parts for adjustment to equality of two comparison surfaces is out of the question. The problem has been dealt with by a committee of the Illuminating Engineering Society,[1] who used a photometer of the form shown in *Fig.* 72. This consists

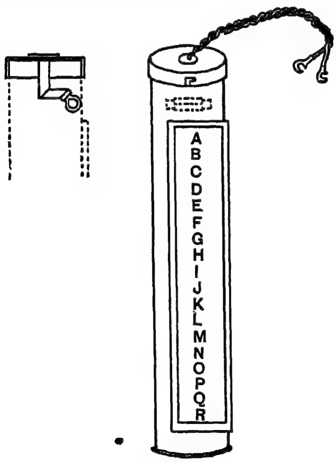


Fig. 72.

of a tube 25 in. long and 3 in. in diameter. The interior is whitened and illuminated by a small electric glow-lamp placed at one end as shown in broken line. A slot extending for nearly the whole length of the tube is covered by a strip of thin metal painted white and having in it perforations in the form of letters of the alphabet. Since the illumination inside the tube decreases gradually from top to bottom, the top letters appear brighter and the bottom letters darker than the surface surrounding them. An intermediate letter can be found at which the illumination of the interior of the tube is the same as that of the outside of the strip. This particular letter becomes practically invisible with a steady light, but with a flickering light the point of balance is continually moving up and down. Letters of the alphabet were chosen because they were monosyllabic and therefore quickly transmitted from the observer who had to watch the photometer continuously, and the assistant who wrote down the observations at close and regular intervals.

§ (123) Candle-power of Flares.—The flares on which tests were made by the committee burned for various intervals of the order of half a minute, and a typical curve showing the variation of candle-power during this time is given in *Fig.* 73. Candle-powers as high as 130,000 were dealt with. The results can be most conveniently expressed, for comparative purposes, in candle-power seconds per gram of composition, a particular example of flare being found to have the value 4000 for this quantity.

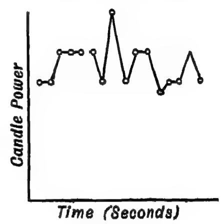


Fig. 73.

X. Measurements of Low Intrinsic Brightness

A third class of photometric measurement demanding special treatment is that in which the brightness of the luminous surface to be measured is very low (below 1 metre-candle). This problem arises, principally, in connection with the study and use of self-luminous compounds, either those depending on photo-luminescence or those in which the exciting agent is some radioactive material, generally radium and its disintegration products.

§ (124) The Colour Problem and Purkinje Effect.—It is impossible by any optical means to increase the brightness of a surface, so that as these luminous compounds, either in bulk or as applied to surfaces in the form of figures or other markings, have luminosities varying from 0·5 to 0·001 metre-candles or even less, the problem of accurate brightness measurement becomes a very difficult one. The difficulty is increased by the fact that the light given by these compounds is often

[1] A. P. Trotter, *Illum. Eng.*, London, 1918, xi. 253.

restricted to one part of the spectrum. Thus, for example, the light given by a compound in which a specially treated zinc sulphide is rendered luminescent by the action of radium is a very decided green (see article on "Radioactivity," § (15)), and as the brightness is below the limit of the Purkinje effect, it is necessary to ensure that the surfaces whose brightnesses are being compared are as nearly as possible of the same colour. It is also especially desirable, in the case of low brightness photometry, to ensure that the dividing line between the surfaces being compared shall be as fine and imperceptible as possible.

In the case of a luminous compound, the problem of photometric measurement requires different treatment according as the brightness of the compound in bulk or that of the markings of the finished design are in question.

§ (125) PHOTOMETRY OF LUMINOUS COMPOUND IN BULK.—In the first case the measurement may be readily made by means of the apparatus shown in *Fig.* 74. S is a sheet of

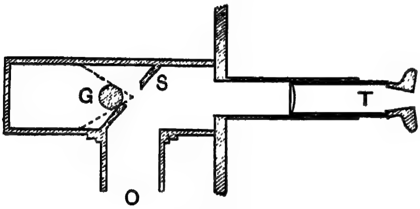


FIG. 74.

matt white celluloid, having at its centre an oblong aperture, so proportioned that when looked at through the telescope T it has the appearance of a small square. This aperture is cut with a sharp bevel to ensure as little separation as possible at the edges. The luminescent compound is contained in a glass tube G, and wedge-shaped guides are arranged inside the box to assist in placing it centrally behind the aperture in S, no matter what the diameter of the tube. The celluloid sheet S is illuminated by a standardised lamp, the light from which, after passing through a green glass or gelatine filter, enters the box at O. The illumination at S is calculated from the candle-power and distance of the lamp, the transmission ratio of the green filter being determined by measurement of the brightness of a given sample of compound by a number of observers with and without the filter in use.

The only filter remaining is that relating the brightness of S, as seen in the direction of the telescope, with its illumination from the direction of O. This may be determined by measuring, with a surface-brightness photometer (see § (64)), the brightness of S as seen in the direction of T, and the brightness of a similar piece of white celluloid placed in the position of S but normal to the incident light. The factor obtained in this way embodies both the reflection ratio of the card under the particular conditions of use, as well as the reduction of illumination according to the cosine law. This form of apparatus is that used at the National Physical Laboratory,[1] but somewhat similar forms of apparatus, designed for the same purpose, have been described by Blok [2] and Clinton.[3] A method in which a circular disc of the material under test placed side by side with a disc of variable brightness has been described by Andrews,[4] while Dorsey [5] places the tube of compound in front of a surface of variable brightness.

§ (126) ADAPTATION OF THE EYE.—In all measurements of low intrinsic brightness the observer's eye must be "dark-adapted" by a lengthened stay (of at least a quarter of an hour, or more if the general illumination to which the eye has been previously exposed be very great) in the photometer-room. Further, as many radioactive luminescent substances are also photo-luminescent, the compound to be measured should be kept away from daylight or strong artificial light for at least half an hour before measurements are made. Luminous compound is generally measured while enclosed in a glass containing vessel. If new, the absorption of the intervening glass wall may be assumed as 9 to 10 per cent, but the action of the rays from the radioactive material is such as to cause a darkening of the glass with lapse of time so that the absorption may increase by as much as 1 per cent per month even when the glass is thin.

§ (127) PHOTOMETRY OF LUMINOUS-PAINTED DIALS.—The measurement of the brightness of the markings on a dial painted with luminous compound is again a special problem in photometry. No constant relation exists between the brightness of the original compound and the luminosity of the markings, and two dials painted with the same compound may have very different luminosities.

The apparatus used at the National Physical Laboratory [6] for the purpose of luminosity measurements of dial markings is shown in *Fig.* 75. L is an electric lamp, the light from which, variable at R, passes through a green filter F and illuminates two screens S, S, each consisting of two thicknesses of thin white paper with a space between them. This

[1] Paterson, Walsh, and Higgins, *Phys. Soc. Proc.*, 1917, xxix. 219.
[2] *Illum. Eng.*, London, 1917, x. 76.
[3] *Ibid.*, 1918, xi. 260.
[4] *Gen. Elect. Rev.*, 1916, xix. 892.
[5] *Wash. Acad. Sci. Journ.*, 1917, vii. 1.
[6] Paterson, Walsh, and Higgins, *loc. cit.*

arrangement ensures perfect diffusion of the light and an even brightness over the whole surface of the second sheet. In front of the

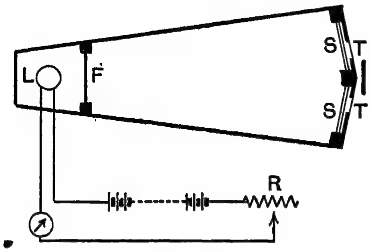


FIG. 75.

screens are placed two stencils T, T, each consisting of a sheet of thin brass with figures or markings cut out on an engraving machine so as to be identical with the luminous markings of the dials under examination. Two comparison dials are used, one on each side of the test dial, so as to overcome errors due to inequalities in sensitivity to green light on different parts of the retina. The instrument is calibrated by measuring, for different values of the current passing through the lamp, the brightness of the surfaces immediately behind T, T, using for this purpose a surface-brightness photometer. The surface brightness of the luminous markings of dials such as those used for aircraft instruments is between 0·2 and 0·05 metre-candles when the compound has been freshly mixed. The luminosity falls off with lapse of time, however, and dials which have been painted for six months or a year may have deteriorated in brightness so much as to be practically useless [1] for their intended purpose.

J. W. T. W.

PHOTOMETRY applied to the estimation of refraction and eyesight, generally by the use of perception scales. See "Ophthalmic Optical Apparatus," § (6).

PHOTOPHONE, RANKINE'S, for transmission of speech by light. See "Sound," § (61).

PHYSICAL PHOTOMETERS: instruments in which physical phenomena are used to measure light, instead of employing visual effect. See "Photometry and Illumination," § (32) *et sqq.*

PIANOFORTE, THE

THE piano belongs to the percussion class of musical instruments in which the sound is produced by giving a sudden impulse to the vibrating system and then leaving it to itself —as contrasted with wind instruments, for instance, in which the vibration is maintained as long as desired by the continuous supply of energy. The sustaining power of percussion instruments is determined by the amount of energy that can be supplied initially and the way in which this energy is dissipated.

The piano is a development from the harp in which the strings were plucked by hand, the next stage being the spinet and harpsichord, in which the wires were plucked by quills. But this method of setting the wires in motion gave none of the light and shade so necessary for the artistic rendering of a musical composition. Hence the complications of the later harpsichord, in some of which by means of stops, as on an organ, the number of wires that were plucked when a key was depressed could be varied; also to some there was added a venetian shutter, which could be opened or closed by a pedal in the same way as the swell box of the organ. All this was done away with and control of the sound was restored to the performer by the substitution of a hammer for the quill, which was first effected by Christofori, an Italian, about 1709. From this original and very primitive instrument the pianoforte has developed by very many steps, generally small, the most important of which have been the inventions of the actions now used in the upright and grand pianos and the introduction of the iron frame. We shall return to these later.

§ (1) THE WIRES.—The original vibrators in the piano are the wires, or *strings* as they are called. There are three to each note from the treble down to about the 32nd note from the bass end (say to E in the bass clef). In long grands they continue another octave. Hence the term *trichord.* Each of these is a simple wire of high tensile steel, left hard from the last one or two drawings so as to give it as high an elasticity as possible. The breaking strain of a good pianoforte wire exceeds 140 tons per sq. in. The nineteen notes from D♯ down to about A in the bass have two covered wires to each note of steel about 1 mm. diameter wrapped with soft copper wire of ·6 mm. diameter. The last twelve notes have only one wire to each note; these are also of steel ranging from about 1·1 mm. to about 1·5 mm. diameter, but wrapped with copper of increasing size from 1·016 mm. diameter to 2 mm. diameter. The lowest notes of all pianos, except of some of the long grands, have two layers of the copper wire, and the total diameter of the wire may be as great as 7 mm.

When a wire is stretched between two fixed supports and set in vibration it will, as is shown in the text-books on sound, execute p vibrations a second, where $p=\sqrt{T/m}/2l$; l being the length of the vibrating segment in cm., T the stretching force in dynes, and

[1] See "Luminous Compounds," § (6) iv.

m the mass in grams of the wire per cm. length. The wire may, however, vibrate not only as a whole—in which case l will be the whole length L of the wire between the two supports—but can also vibrate in 2, 3, 4, or more segments, and then l will become L/2, L/3, L/4. Thus (neglecting rigidity) the same wire may be caused to give any one of a series of notes of which the frequencies are p, $2p$, $3p$, $4p$, p being the frequency of the wire when it vibrates as a whole. In all ordinary cases when a wire is plucked or struck with a hammer it does actually move in such a way as to be equivalent to the superposed effect of all these frequencies. Its equation at any time t is thus given by $y=\sum_1^n A_n \sin(2\pi nt/\mathfrak{T}) \sin(\pi nx/L)$, $\mathfrak{T}$ being the periodic time of the fundamental; the quality of the sound it will produce depends upon the relative amplitudes of the component terms in this series.

The maximum energy that can be given to a wire is dependent on the mass of the wire set into vibration; it can therefore be increased by using larger and longer wires. The three wires of each note are tuned to unison and struck simultaneously by the hammer; this is obviously a device for multiplying the energy which can be given to the wires and by them dispersed *via* the soundboard as a musical sound. The introduction of the iron frame, by making it possible to increase the tension on the wires, and therefore to use longer and heavier ones, is one of the chief factors which distinguishes the modern piano from that of say fifty years ago. But the increase in strength of the frame would have been useless unless the tensile strength of the wire had been increased to correspond, and if this strength can be still further increased it may be possible to get a still greater sustaining power.

The series above given only represents the motion of the wire to a first approximation, for the rigidity of the wires influences the pitch of the higher notes and of the upper partials of all the notes, causing the frequency to be higher than that given by the formula; for the restoring force due to the tension of the wire is augmented by that due to its stiffness, and although the latter has little effect on the long vibrating segments of the lower partials, it obviously will play an increasingly important rôle as these segments become shorter. Thus the higher partials of any string in vibration will be sharpened. The effect on the pitch is the same as if the wire were perfectly flexible and the stretching force increased by $\pi^3 r^4 Y/4l^2$ dynes, where r is the radius of the wire, l the length of the vibrating segment, and Y is Young's modulus of elasticity. This sharpening is not in general noticeable, for the amplitudes of the partials after the third or fourth is very small, and they can only be heard with difficulty. But should the cross-section of a wire be an ellipse instead of a circle, the rise in pitch due to the rigidity will be greater for vibrations executed in the plane containing the major axis of the ellipse than for those executed in the plane at right angles to this. Thus, as usually the wire will not be vibrating wholly in one of these planes, each partial will yield two frequencies differing slightly from one another, which will therefore beat with one another. This beating is easily heard, although the partials which give rise to it would otherwise have been too weak to be perceived, and, as it makes the wire sound out of tune, such a wire is described as "false." The only remedy is the replacement of the wire by another. It can thus be seen that the perfect circularity and uniformity of a piano wire is of great importance.

It is evident from the formula that as the frequency depends on the three factors—length, tension, and mass—any two of the three may be taken as independent variables and be chosen arbitrarily, and then the third can be determined to obtain the pitch required. In practice the tension is made uniform for all the strings of a piano; for pianos in which the tension is the same throughout have been found to keep in tune best, no doubt because with a uniformly distributed load, temperature variations affect the instrument as a whole and do not throw the notes out of tune with one another. The tension adopted by various makers range from 160 to 190 lbs. on each string. This is well below the limit of elasticity of all but the smallest gauge wires, as can be seen from the following table which has been extracted from figures given in Wolfenden's *Art of Piano Construction.*

Note and Number.	Length in Mm.	Diameter of Wire in Mm.	Breaking Strain in Lbs.	Actual Tension in Lbs.
C 88	5·40	·800	275	173
C 76	10·20	·850	306	173
C 64	19·25	·900	332	173
C 52	36·40	·950	366	173
C 40	68·80	1·000	384	173
C 28	131	1·050	410	173

The tension of the intermediate notes varies from 173 lbs., for the same gauge wire is always used for several successive notes, and as the lengths are increased in a regular manner, the sudden variation in the sizes of the wire has to be compensated by corresponding changes in the tension.

It might at first sight seem possible to keep to one size wire throughout and to use the length only as the factor to obtain the change in pitch. The wire for any note would then be double the length of the wire belonging to the octave above. This is nearly done for the upper octaves, but it must not be forgotten that rigidity plays an important part, and if this relation were adhered to, it would make the top notes too short and too stiff, and those of the lower octaves would be too limp as well as unmanageably long. The length of the wire is therefore graduated to the extent of reducing the ratio to 17 : 9 for all but the lower octaves. In order to allow of the use of wires of lengths following this ratio as far down towards the base as possible, the practice has arisen of "overstringing," that is to say, the bass wires are carried more or less diagonally across the tenor wires, the latter wires sloping some 20° to the vertical downwards to the left, while the former slope an equal or greater amount to the right. The change from the last wire sloping one way to the first sloping the other is called a "break," and as these wires pass over different and widely separated bridges on the soundboard, there is liable to be a serious difference between the sound yielded by those two notes, and it requires judgment and skill to minimise this. Other changes are for this reason usually avoided at this point; for instance, the change from the simple steel wire to the covered wire is often made four or five notes above this point.

Even with the extra length that overstringing gives, the lower octaves cannot continue to increase in length in the same ratio, in fact the wires of the last octave are practically all one length, the change in pitch being produced by the additional thickness of the copper wire which is wrapped round the central steel core. Sometimes in cheap pianos the makers lower the tension in this octave to reduce the weight of the costly heavy copper wire. In any case, except in the long concert grand, the resultant quality of the octave is usually poor; but these notes are of comparatively small importance as they are much less frequently used prominently, and are generally accompanied by the octave above, which helps to mask their poor quality.

Strips of felt are threaded in and out between the wires above the top bridge, and between the bottom bridge and the hitch-pin, to prevent these portions of the wire from vibrating; one or two makers have endeavoured to make the lower end of the wire equal in length to the part between the bridges, in the hope that this would form an additional resonator. It is difficult to see how such a wire can add to the initial energy, but it might help to avoid a useless dissipation of energy in the felt just referrred to.

§ (2) The Soundboard.—Although the wires are the original vibrators in the piano, their vibrations are themselves to all intents and purposes inaudible; for as the wire swings to and fro, the air in front of the moving wire slips round to fill the space left behind it by its motion, and no compressions and rarefactions of appreciable amplitude are produced in the air. To a slight extent the pressure of the wire on the fixed upper bridge transmits the vibration to the iron frame, thence it passes to the case of the instrument and the floor of the room, and these in turn transmit it to the air. But in the main the transmission of the vibrations of the wire to the air is by the medium of the soundboard, this thus plays almost as important a part in the speaking quality of a piano as does the body of the violin in the tone yielded by that instrument. The *soundboard* of the piano (as of the violin) is made of a wood in which the ratio of the elasticity to the density is as high as possible. Pine fulfils this best, and "belly-wood," as it is called, is usually Norway spruce, *Picea excelsa*, or sometimes *Abies pectinata*. The best trees are those growing in high altitudes on a mountain side where the growth is slow and regular. Before the war the best came from the Roumanian forests; some is now being obtained from British Columbia. The wood is cut on the quarter (*i.e.* radially) into strips 7 to 10 cm. wide and about 1 cm. thick, and these are carefully jointed up to form a more or less rectangular board with the grain running approximately parallel to the long bridge, that is from the top right-hand corner to the bottom left-hand corner at an angle of about 40° to the horizontal. This board is stiffened at the back by about ten parallel wooden bars, 2 to 3 cm. square, usually tapered off at the ends, which are glued across the soundboard at right angles to the grain of the latter. In gluing them on the makers attach great importance to "bucking" the soundboard—that is to say, making the front surface convex. The centre should be about 1 cm. above the edge. This curvature is produced in one or more of three ways:

(1) By planing the bars to a curve so that when they are glued on they shall tend to draw the board round to their own curvature.

(2) By the use of a concave table on which the board is laid while the bars are being glued on, and into which it is forced by the pressure of a large number of *go-bars* which are sprung in between the bars and the roof, as

shown in *Fig.* 1, where BB is the hollow-top table on which the soundboard CC is placed; AA is a firm wooden roof; DD are the bars to be glued on the back of the soundboard; GG, G′G′ are the *go-bars* which are sprung in between DD and the roof; E is a block of wood to prevent the go-bar from denting the bar.

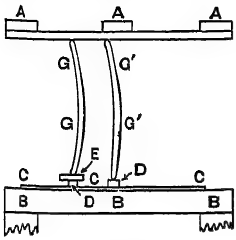


Fig. 1.

(3) By heating both the bars and the soundboard for some time before gluing. This causes the bars to expand lengthwise, and (by drying it) causes a contraction of the soundboard across the grain. So when the whole is cold and it has reabsorbed a little moisture it tends to curve.

This curvature helps the soundboard to sustain the pressure which is caused by the *down-bearing*, or the resolved part normal to the surface of the soundboard of the tensions in the wire; for in order to keep the wire well down on the bridge during its vibration the top of the bridge is left at such a height that it raises the wire a little (2 to 5 mm.) above its natural position. The pressure of the wires forces the bridge down, but it is considered fatal if it should be forced down so much that the soundboard becomes concave. The down-bearing is usually measured by the angle the two segments of the wire make with one another. This may be as great as 1½°.

The edges of the soundboard are fixed very firmly by hard wood fillets to the wooden framework or *back* of the piano, which back also carries the iron frame and is the foundation of the case of the instrument. Any want of rigidity round the edges of the soundboard will not only lead to a dissipation of energy in friction which should have been given out as sound, but may also cause noises which are difficult to locate and eliminate.

By placing the piano with its soundboard horizontal, scattering sand upon it, and then striking a string with a hammer, nodal lines are often obtained. This shows that to some extent stationary waves are formed, and that the soundboard does not vibrate as a whole, but that some parts of it are in a different phase to the rest. This partly explains a fact which at first sight seems curious, namely, that it has been found advantageous to limit the size of the soundboard. Usually the lower right-hand corner, and sometimes also the top left-hand corner, are either cut away or are prevented from vibrating by a heavy wooden bar glued on behind called a *dumb-bar*.

Just as the varnish on a fiddle is supposed to be of great importance, so some have held that the varnishing of the soundboard is also of great importance. But it seems doubtful if the varnish plays any other part than that of preserving it from damp.

Most makers seem to have had at some time the idea that by imitating the violin and using a double soundboard, with or without apertures, it would be possible to improve the piano; but all such attempts have been failures.

In order to give the greatest possible length to the bass strings, and at the same time to avoid the want of response which would be caused by placing the bridge close to the edge of the soundboard, the short bridge on which the overstrung wires are carried is usually a *floating bridge*. The meaning of this is shown in the diagram, where B is the fillet at the edge of the soundboard A, W the overstrung wire which passes over the bridge C and

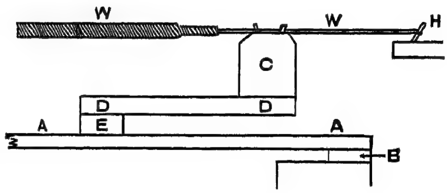


Fig. 2.

terminates at the hitchpin H. The bridge C, instead of being glued directly on the soundboard A, is carried on a thin board D, which is glued to a block E attached to the soundboard at some distance from its edges. This device enables the vibrations to be communicated to a flexible part of the soundboard and greatly improves the quality of the tone produced.

§ (3) The Hammer.—The oldest hammers were of wood covered with leather. The old hand-covered hammers were of wood covered with a number of layers, of which the inner ones were hard felt and the outer ones softer felt. The modern "machine-covered" hammer has only one layer of felt over the wood. The whole set for a piano is covered in one operation. The felt when first made is 10 to 15 cm. thick, but it is shrunk by the makers until it is about 2 cm. thick at the bass end, 3 to 4 mm. thick at the other. It is then cut to a roof shape as shown at BB, *Fig.* 3. It is glued and bent up round the row of hammers AA, being pressed upon them by hot metal moulds in a machine. Any degree of hardness required can be obtained by suitably regulating the pressure on the

moulds. When the glue is dry, the hammers are cut apart with a sharp knife. In this way a very regular gradation of weight, thickness, and elasticity from bass to treble can be obtained. Moreover, the density of the felt in any hammer is greatest in the centre of the felt next the wood, and gets less as the outer surface is approached. This is a very important matter, for it determines the relative intensities of the fundamental and the upper partials, or in other words, it determines the quality of the sound produced.

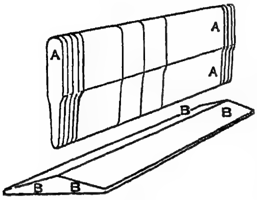


Fig. 3.

To understand the action of the hammer, stretch a wire thirty or more feet long with a small force—say a pound weight—gently pluck it near one end, and put a finger lightly upon it. The plucking will produce a pulse which will travel to and fro along the wire and can easily be felt each time it arrives at the finger. If it be struck with a rod a sort of momentary dent will be formed which will travel away along the wire in both directions from the point struck. When it reaches the end the dent will be inverted and return as a hump. In so doing it is evident that it will give the support an impulse, just as a man who is exercising with a pair of heavy dumb-bells can shake the floor by bringing them down or up suddenly. It is this reversal of the pulse in the wire each time it arrives at the bridge that is mainly responsible for the communication of the vibration of the wire to the soundboard. To return to our experiment, it is evident that while a smart blow with a metal rod will cause a sharp dent, a blow with a soft felt hammer (such as a bass drumstick) will cause a smooth dent of much less curvature. If the hammer were of such weight and consistency that the blow lasted for half the natural period of the stretched wire, while its strength gradually increased to a maximum and then died gradually away as the hammer rebounded, the pulse produced would have a length equal to the length of the wire; thus an analysis into the Fourrier series to which it is equivalent and which expresses the motion, would result in a series in which the amplitude of the fundamental was predominant; a similar analysis of the pulse produced by a metal rod would require a larger number of terms to represent it with any closeness. Thus if the sound produced when the piano wire is struck by the hammer is to be largely composed of the fundamental and the first one or two partials, the hammer must have suitable weight and elasticity. It should strike a blow which commences gently, rises to a maximum in a time which will vary with the pitch of the note, and then dies away as regularly. A moment's thought will show that if the outer layer be soft and curved so that it touches first at a point, the blow will commence gently, then as the blow proceeds, if the deeper layers are harder, the pressure will increase rapidly attaining a maximum at the moment of greatest compression of the felt when the hammer and wire are relatively at rest; after this as the hammer rebounds the pressure will die away. As the coefficient of restitution of felt is not very high, the force will be less during the rebound and the pressure-time curve will not be quite symmetrical.

The point of impact of the hammer on the string is of great importance. From about the 40th note from the bass end downwards to the bass end the string is struck at about one-eighth of the length of the speaking part of the string (that is the distance between the two bridges) from the upper bridge. From this note upwards towards the treble, the fraction of the length from the top bridge gradually changes until at the top it becomes one-fourteenth or even one-sixteenth of the speaking length. Helmholtz showed that when a string was struck at any given point every partial which had a node at that point should be absent, and he thought that by striking the string at one-seventh or one-ninth of the distance from the end these partials which are inharmonic would be eliminated; but it is obvious that this cannot be the controlling factor, since at one-eighth the partials eliminated would be in tune, and this is the fraction chosen for all those notes for which such partials would be most obvious. Now it is a fact that the amplitudes of the partials above the third and fourth in a good piano are relatively small, and though their existence can be proved by means of resonators, they are almost inaudible without such aid and are probably negligible. Thus it is much more probable that the point found most suitable is really determined by the motion of the hammer. The makers say that the treble notes become more brilliant when the wires are struck nearer the end. Towards the top notes of the piano the *thud* produced by the blow of the hammer is audible if the point of contact is not nearer the end of the string than one-eighth of its length, and this may, therefore, be another reason for the point of contact chosen for those notes.

§ (4) The Action.—The action now universally adopted for all upright pianos is developed from the action patented by an Englishman named Wornum in 1826.

Wornum's action is virtually an adaptation of an earlier one of Backer's for a horizontal piano. It was very nearly the same as the action illustrated, except that he continued the use of an inclined plane to throw the jack out of the notch of the hammer-butt. The bell-crank lever was first embodied by Erard in his grand action, and was not used in the upright action till much later. Wornum found it impossible to induce the English makers to adopt his action, and it was the French who took it up, so that it became known as the *French action.* Most English makers continued to use the old "sticker" action until nearly the end of the nineteenth century—long after Wornum's action had been adopted universally abroad. A good action (1) must give the performer full control of the blow, so that he can vary the intensity of the sound produced with ease and certainty; (2) must withdraw the hammer from the string immediately—even if the performer hold the note down—or the motion of the string will be damped by the hammer; (3) must not allow the hammer to bounce back and hit the string twice, unless a second movement is imparted to the key; (4) must enable the performer to obtain rapid repetition of the blow when desired, and (5) this repetition should be obtainable without waiting for the key to rise to its initial position. All this the Wornum action provides. When the front of the key K is pressed down the pilot P, which is screwed into the back end of the key, presses against the rocker A, which is pivoted at T. To the rocker is pivoted the jack J. The top of the long end of the jack rests in a notch in the butt B or hinge of the hammer, and thus pushes this up, causing the hammer to move towards the string FF′. Just before the hammer strikes the string, however, the rocker A brings the toe of the jack up against the set-off Q, and this causes the jack to rotate clockwise, and so lifts its top forward out of the notch of the butt B, leaving the latter free. This should happen when the hammer is 2 to 3 mm. from the wire, so that it is only the inertia of the hammer which causes it to strike the wire. It will then rebound and would fall back on the rail R were it not for the check C, which comes against E, which has been raised by the motion of A to which it is attached. In this way the hammer is held up a short distance from the string as long as E is held against the check C, *i.e.* as long as the performer keeps the key down. This has two important results: (*a*) it catches the hammer and prevents it bouncing back to strike the string a second time; (*b*) it holds the butt up so that as the key is raised the jack can fall back into the notch almost as soon as it is free of the set-off, and this enables the performer to cause the hammer to strike again without having first to let the key rise up to the top, as he would otherwise have to do. The damper D is carried on a wire inserted in a wooden lever hinged at M, and the lower end of this lever is pressed to the left by the spoon S when the rocker A is rotated.

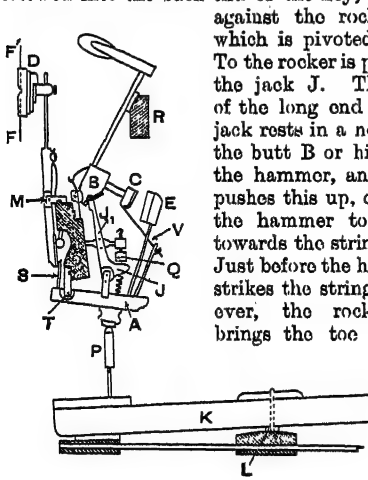


FIG. 4.

The hinges or *centres*, as they are called, of this action are made by brass wires passing through a hole bushed with thin cloth. This yields a smooth working hinge free from back-lash and noise. The bearing surfaces of the several parts of the action are covered with soft fabrics to avoid noise. The top of the notch in B into which the jack falls is lined with doe-skin, as it has to transmit the blow, but the back of the notch is lined with soft felt, as it merely has to receive the end of the jack and stop it as noiselessly as possible.

An interesting point arises in regard to the best position for the interface between the pilot P and the foot of the rocker A. In order to reduce wear at this surface it is obvious that there should be no sliding motion between P and A, and that it should be a pure rolling one if possible. Now P is rotating about the fulcrum of the key K, while the rocker rotates about the centre T, so that to produce a rolling motion the point of contact between P and A should lie throughout the motion on the line joining T and L. This determines the depth of the foot of the rocker. (It should also determine its curvature, but this is usually ignored.) The same consideration applies to the point of contact between M and S which should lie on the line joining the centres M and T; also to the contact between J and Q which should lie on the line joining T and the centre or hinge of the jack.

An important feature of this action is the tie V, which has two objects: firstly, as the hammer is only about 3° from the vertical when it strikes the wire its weight cannot help it to come away after the blow, and if the blow were a gentle one it might be too slow in returning after the blow; and secondly, the jack might get below the butt instead of merely falling into the notch. Hence Wornum provided the string to connect the hammer

block with the rocker, and so to prevent any such undue separation.

§ (5) THE GRAND ACTION.—This beautiful action was invented by Sebastian Erard in 1827 and is now used almost universally. The underlying principles are the same as those of the upright action, already described, but the mechanism by which the results are achieved was quite original. *Fig.* 5 shows a modern form of this action which is almost identical with the action as Erard left it. The wippen 2

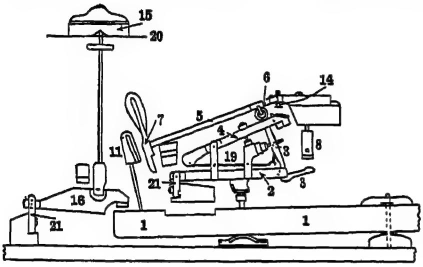


FIG. 5.

which is pivoted at its rear end to the flange 21 carries the jack 3, which presses against the roller 6 attached to the hammer shank 5, and so causes the hammer head 7 to strike the string 20 when the key 1 raises the wippen 2. Just before this happens, the toe of the jack 3 comes into contact with the button 8, and this rotates the jack clockwise and removes it from below the roller 6; so that the hammer is quite free when it strikes the wire. After the blow it falls back upon the check 11. The repetition lever 4, which is pivoted near its middle to an upright fixed on the wippen 2 (and therefore moves with it), assists the check in holding up the hammer for a repetition of the blow. This it does because its front extremity comes into contact with the rail 14 before the jack reaches the button 8, and being held up by spring 19 it remains in contact until after the jack has left the button. The damper 15 rests on the top of the string, and is lifted by the lever 16, which is pivoted to the flange 21 and lifted by the far end of the key 1.

R. S. C.

PIANOFORTE: a stringed musical instrument in which the sounds are excited by felt-faced hammers striking the strings and then leaving them to vibrate. See "Sound," § (29); also previous Article.

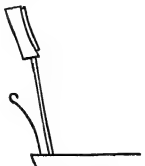
PIANO-PLAYER, THE

§ (1) DEVELOPMENT.—The piano-player has developed from the old barrel-organ, in which the music to be played was produced by opening the pallets by stiff wires set in a wooden or metal barrel. A patent was taken out in 1694 for a machine for playing on any keyed instrument, including the virginal and harpsichord, so that mechanical players are of some antiquity. The pierced paper rolls appear to have first been used for wind instruments. The openings in the rolls were passed over other openings communicating with chambers in which harmonium reeds were placed. The air was able to pass through the reed and cause it to sound when the opening in the paper came over a hole. The paper rolls evidently could not be used to strike a note on a piano directly, but it is obvious that the admission of the air through a cut in the paper could be used to operate a bellows to strike a note. In 1863 Fourneaux in France took out a patent for striking a note on a piano by the aid of air; but it was Gally of New York in 1881, Bishop and Down in England in 1885, and Kuster of New York in 1887 that really developed the pneumatic player. The player was sold as an independent instrument up to 1902, and even after that date many separate players were sold. Gradually, however, its incorporation into the piano itself has become universal.

§ (2) PNEUMATIC ACTION.—The main principles of the pneumatic action had been developed by the pipe-organ builders; it was the provision of means for varying the force of the blow, traversing the music roll over the tracker-bar, adjusting and varying the tempo, and emphasising the melody, which required to be developed in order to make the player a satisfactory instrument.

The mode of action of the player is shown in its simplest form in the accompanying figure (1). A long sheet of thick perforated paper 2 travels from the music roll 3 to the take-up spool 1 over the tracker-bar 4. This bar is pierced with a row of 88 narrow slots about 3 mm. apart, each hole corresponding with one note on the piano which it is to actuate. When a perforation in the paper passes over one of these slots, it allows a little air to enter the duct 5, which communicates by the tube 6 with the primary chamber 7. This chamber is divided into two parts by the thin flexible circular diaphragm 8. The upper part of this chamber 9 is maintained partially exhausted of air, so the sudden difference of pressure between the upper and lower parts of this chamber lifts the diaphragm. Just above this diaphragm is the lower end or disc of the dumb-bell-shaped primary valve, which is therefore also lifted. The movement of the valve closes the communication between the valve tube 12 and the exhausted primary chamber 9, and puts it into connection with the atmosphere by the lifting of the upper end 11 of the valve. Air is now able to pass through the valve-tube 12 into one side, 14*a*, of the secondary pneumatic chamber 14. This chamber is also divided into two parts by

the secondary diaphragm 13; the part 14*b* is kept partially exhausted of air, and thus, when air is admitted to the other side of the diaphragm, the difference of pressure causes it to move, carrying with it the secondary valve 15, 16 to which the diaphragm is joined. This movement closes the opening from the open air to the channel 17, and puts this channel



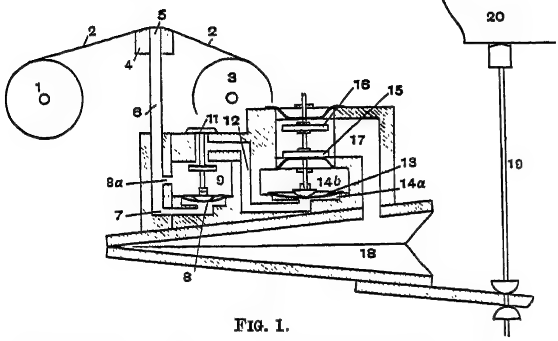


FIG. 1.

into connection with the exhausted chamber 14*b*, so that air is withdrawn from the pneumatic 18, causing it to collapse, and to lift the action 20 by means of the connecting rod 19. There is a tiny hole 8*a* called a bleed-hole connecting the chamber 9 with the tube 6; so that when the perforation has passed across the aperture in the tracker-bar and the slot is once more closed the pressure becomes equalised on both sides of the diaphragm 8. This then falls and allows the primary valve to fall also. The air is therefore sucked out of the tube 12, the valve 13 is drawn back, and the pneumatic 18 is allowed to open, ready to give another blow. All this is very similar to the operation of the pneumatic action in the larger pipe-organs.

§ (3) BELLOWS AND RESERVOIR.—The two chambers 9 and 14 of *Fig.* 1 are kept ex-

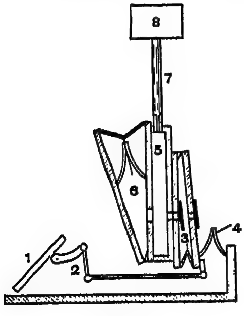


FIG. 2.

hausted by being connected through the action-box 8 (*Fig.* 2), trunk 7, and the bellows-board 5 to the reservoir 6.

The reservoir is a bellows kept open by a spring; the air has to be continuously removed from it by working the foot-pedals 1, which through the pedal-bars 2 force open the bellows 3 against the pressure of the spring 4. The pneumatics of a pipe-organ are worked by air, which is at a pressure of about 30 cm. of water above that of the atmosphere; but while it is obvious that both primary and secondary pneumatics might be arranged to work with an air pressure which is either above or below the atmospheric pressure, the latter is now always used in players, as the reduced pressure holds the music roll down against the tracker-bar, and leakage is avoided.

§ (4) MOTOR.—To carry the music roll from the roll 3 on to the take-up spool 1 of *Fig.* 1, the latter is rotated by a pneumatic motor, which is also driven by air passing to the reservoir 6. The motor consists of five simple bellows

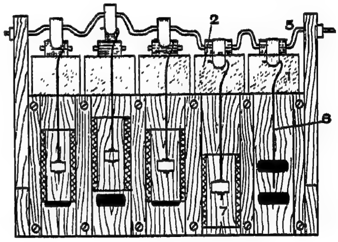


FIG. 3.

1, 2, etc. (*Figs.* 3 and 4), each joined by a connecting rod 4 to a five-throw crank-shaft 5, the cranks being at 72° to one another. The same shaft works slide-valves 7 by arms 6, which alternately admit air to each of the bellows and put it into communication with the reservoir through the exhaust 8.

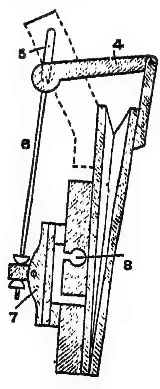


FIG. 4.

§ (5) EXPRESSION.—The player as it has so far been described would run at a constant speed, strike every note with the same force, and altogether fail to pick out a melody. Thus although its execution might be perfect, it would give none of the light and shade and no variety of tempo (except such as might be incorporated

into the original cutting of the original music roll). In fact, it would, artistically, be a failure; for although the soft pedal in the modern piano throws the action forward and so shortens the blow, this alone would be insufficient. A much greater range of power can be produced by the rate at which the bellows are worked; but the change of air pressure so produced would, of course, alter the speed as well as the blow. Some means of adjusting the air pressure of the motor independently, and also that of the chamber 14 (of *Fig.* 1) of the secondary pneumatic which actuates the final pneumatic 18, is therefore essential. This is achieved by the pressure-control valve of which the action is illustrated in *Fig.* 5. There is a chamber 2 communicating both with a small bellows 3 and with the reservoir through the rectangular aperture 1. A spring connects the top of the bellows with a lever 6. The top of the bellows is also connected with a knife-valve 5

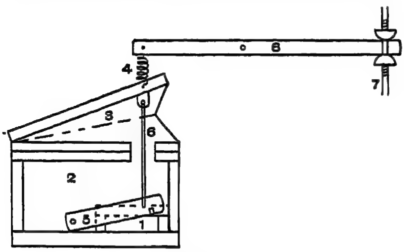


FIG. 5.

by the rod 6. As the pressure diminishes in 2 the pressure of the atmosphere on the outside of the bellows causes it to collapse and to close the valve 5, so stopping the exit of the air. Thus the pressure of the air in 2 is maintained constant; the actual pressure depends upon the position of the lever 6, which latter is joined by the rod 7 to a lever under the control of the operator—generally at the front of the keyboard. One of these valves is put between the motor and the reservoir. It controls the tempo at which the piece is to be played and is set at the beginning of each roll. A second one is put in series with this to enable the operator to produce accelerations, or retardations, or even pauses, in the piece at pleasure. One or more similar valves control the supply of air to the final pneumatic 18 (*Fig.* 1). Frequently there is one to affect all the upper notes on the piano and another to operate upon the lower notes.

§ (6) SOLO DEVICE.—For playing some chords louder than others, or a melody louder than the accompaniment, there are numerous devices. Only two can be described here. In one of these the note or notes to be emphasised are cut a little behind the other notes that should be played together, as shown in *Fig.* 6, where 5, 6, are to be played quietly, but 2, 3, 4, are to be emphasised. The slots of the latter being a little behind those of the former, the corresponding pneumatics will act a fraction of a second later. At the edge of the music roll there is an extra perforation 1, which

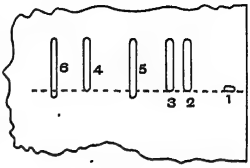


FIG. 6.

actuates a special pneumatic whose function it is to change over the secondary pneumatic chamber 14 (*Fig.* 1) from communication with one of the control-valves just described (*Fig.* 4) set at a lower exhaustion, to a second valve set at a high exhaustion. Thus after the blows have been given by 5 and 6 with a small pressure difference, the other notes are struck with a greater force. The pneumatics for the louder notes will be actuated a little later than the rest. Though this later speech of the emphasised notes might not necessarily be disagreeable, it is possible that the notes are actually struck almost simultaneously, for although the hammers of these notes will start an instant later, yet as they are given a greater velocity they may reach the wires and strike the notes at the same time as the quieter ones.

The second method makes use of a group of perforations at the side of the tracker-bar as shown in *Fig.* 7. The perforations of the

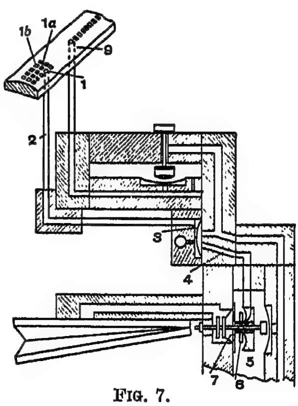


FIG. 7.

notes themselves which are to be emphasised are not altered in any way. Each of the side perforations 1 admits air through a tube 2 to a primary pneumatic 3, which in the usual way actuates through 4 a secondary diaphragm 5. This diaphragm lets a plate 6 come down upon the opening 7 and closes it. The plate

is pierced with a small hole only, and when it closes the opening 7 it obstructs the passage of the air, and only allows the pneumatic 8 (which will be actuated if a perforation arrives over the opening 9 in the tracker-bar) to close gently. Thus if the aperture 9 is opened while 1 is closed, the note is struck gently, but if 1 is open the note is struck loudly. It might be thought that as the perforation which is to open 1 has to pass over the other holes 1*a*, 1*b*, it might cause other notes to be louder than those desired; but if the distance between 1 and 1*a* is less than the time interval between two notes on the roll, this cannot happen. For it will not affect the note if 1 is open either before or after its note is struck, it is only operative if 9 is opened while 1 is open. It is obvious that this is a more complicated way of producing emphasis, but it is a positive method. To some extent it does involve taking liberties with the timing of the notes, in order occasionally to avoid emphasising a note that does happen to require a perforation that would have acted just when the side perforation was passing over its solo hole.

§ (7) Tracker Device. — It is of course essential that the paper roll shall register correctly with the tracker-bar, so that the slots cut in the music roll may pass exactly over the holes they are intended to open. To accomplish this, two holes (1 and 2, *Fig.* 8)

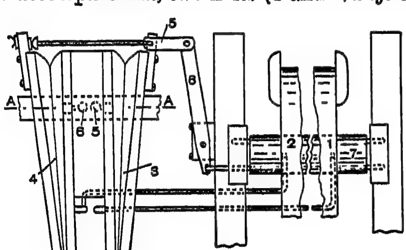


Fig. 8.

are added to the bar; they are close to the edge of the roll and are covered by the roll when it is in its correct position. The holes are connected with the bellows 3 and 4 by tubes as shown. The bellows are normally kept exhausted by small bleed-holes (5 and 6) connected to the air reservoir. When the paper does not go quite straight, it opens one of the holes, 1 or 2, and the corresponding bellows begins to close slowly. By means of the bar 5 (which joins the bellows together, and so has been balancing their pulls) and the lever 6 the tracker-bar is caused to slide along, until it once more registers with the music roll.

§ (8) Electric Players.—These have also been made, a brush making contact with a metallic plate through the perforations. In the earlier machines the current energised an electromagnet which acted directly upon the piano action. This was very unsatisfactory, as if it was attempted to reduce the blow by reducing the current the armature might not be moved at all, on account of the rapid decrease of attractive force with the distance. In an interesting player a metal cylinder is kept in continual rotation. The armatures are arranged a little below and in front of this cylinder. Loose straps pass from the armature over the cylinder to the piano action. When the armature is energised the strap is gently tightened, and the friction of the revolving cylinder rubbing against the strap lifts the action and strikes the blow. Power is of course supplied by motors, which can be controlled either by varying the current or by brakes.

§ (9) Automatic Recording. — Electricity has been very successfully applied to recording the actual performance of a pianist. Contacts on the piano keys are connected to electromagnets in a punching machine, so that the exact instant when the note is struck, the length of time that the key is held down, the use of the pedals, and the minutest variations in tempo are all recorded. In reproducing such a record the operator has therefore only to concern himself with the intensity of the blow.

Many records have the music printed upon them, so that the operator can render the expression more intelligently. This is especially necessary in records which are intended to be used as accompaniments to the voice or violin.

R. S. C.

Piccolo: a wood-wind musical instrument. See "Sound," § (34).

Pipe, Conical, calculation of frequencies of vibration of. See "Sound," § (52) (v.).

Pipe, Open, calculation of frequencies of vibration of. See "Sound," § (52) (ii.).
Correction for mouth and open end in calculation of frequencies of vibration of. See *ibid.* § (52) (iii.).

Pipe, Stopped, calculation of frequencies of vibration of. See Sound, § (52) (iv.).

Pitch of Musical Instruments, effect of change of temperature on, tabulated. See "Sound," § (52) (vii.), Table X.

Pitch of a Musical Note: a term used in music to denote the property of the note which determines its position in the musical range; the number of vibrations per second, or *frequency*, of the sound is adopted as a precise physical measure of the pitch. See "Sound," § (1).

PITCHES, TYPICAL MUSICAL, IN CHRONOLOGICAL ORDER, tabulated. See "Sound," § (7), Table III.

PITCHBLENDE, occurrence of, as ore of radium. See "Radium," § (2).

PLANCK'S RADIATION FORMULA. See "Radiation Theory," § (6).

PLANE-TABLES AND ACCESSORIES. See "Surveying and Surveying Instruments," § (17).

PLANE-TABLES AND METHOD OF PLANE-TABLING. See "Surveying and Surveying Instruments," §§ (9) (iv.), (16).

PLATE-GLASS, MANUFACTURE OF. See "Glass, § (18) (v.).

PLEOCHROISM. See "Polarised Light and its Applications," § (17) (i.).

POISSON'S RATIO: the ratio of the lateral contraction to the longitudinal extension of an elastic material; its value in glass is 0·21–0·28. See "Glass," § (26) (iv.).

POLAR DIAGRAMS, as used in photometry: diagrams showing graphically the candle-power of a light source in various directions. See "Photometry and Illumination," § (37).

POLARIMETRY

§ (1) DESCRIPTIVE AND HISTORICAL.—When certain crystals and liquids are placed between crossed Nicol prisms, the light will not in general remain extinguished. If the light source is homogeneous, or approximately so, by rotating the analyser (*i.e.* the polarising prism nearer to the eye, the farther one being called the polariser) a new position is found where complete extinction is again obtained. This can be explained by regarding the crystal or liquid as having rotated the plane of polarisation of the light as it passed through the substance.

The phenomenon can be illustrated schematically by means of *Fig.* 1. Consider the case

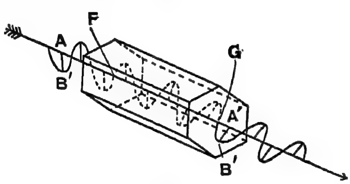


FIG. 1.

of a plane polarised beam of light passing through a crystal of quartz along its optic axis, the beam before entering the quartz being polarised in the azimuth perpendicular to AB. The plane of vibration is in the plane of AB. When the light enters the quartz, this plane is gradually orientated by an amount that is proportional to the thickness of quartz traversed, until it emerges at G, with the direction of the vibration now lying in the plane A′B′. Now that the light has left the "optically active" quartz, the plane of vibration of the farther course of the beam will be that of A′B′; and the plane of polarisation will be rotated, by passage through the quartz, by an angle equal to that between AB and A′B′.

Arago [1] observed in 1811 that the effect produced when a plate of quartz cut perpendicularly to its axis is placed between a polariser and analyser differs from that caused by a plate of mica in the same position.

Biot [2] was the first, however, to point out, in 1812, that a rotation of some kind was involved; he showed [3] also, in 1915, that this property was exhibited by turpentine and a number of other substances. The rotation depends on the thickness of active medium traversed and on the wave-length of the light being inversely proportional (approximately) to the square of the wave-length. The name Rotatory Dispersion has been given to this property.

Biot's polarimeter [4] had as polariser a plate of black glass, set at the polarising angle, viz. the angle at which polarisation was most complete. This was connected with a fixed divided circle by means of a V-shaped trough in which tubes containing liquids could be placed. The "analyser" consisted of a double-image prism, a prism, that is, giving two images, one formed by the "ordinary" and one by the "extraordinary" beams, these being polarised in planes perpendicular to each other. This prism was mounted in a tube carrying an index arm by which the angular rotation of the analyser could be read on the divided circle. The analyser was rotated until the image of the light source formed by the extraordinary beam was most near to extinction. When the optically active substance was placed in the trough, the image (from the extraordinary beam) would again appear, and the analyser rotated until extinction was obtained once more. It was with this very simple apparatus that Biot made the elementary observations that are of fundamental importance in polarimetry.

Ventzke [5] shortly afterwards employed Nicol prisms as polariser and analyser, thus securing *complete* polarisation at the polarising end, and the elimination of the double image at the analyser.

[1] Arago, *Mém. de la prem. classe de l'Inst.*, 1811, xii. 93.

[2] Biot, *Mém. de la prem. classe de l'Inst.*, 1912, Part I. pp. 262-264.

[3] Biot, *Bull. Soc. Philomat.*, 1815, p. 190, or *Ann. de Chim. et de Phys.*, 1817, iv. 90.

[4] Biot, *Compt. Rend.*, 1840, xi. 413; *Ann. de Chim. et de Phys.*, 1840 (2), lxxiv. 401.

[5] Ventzke, *Erdman's Journ. für praktische Chemie*, 1842, xxv. 65.

Mitscherlich[1] was the first to appreciate the importance of monochromatic light in polarimetry; he also improved the illumination by placing a convex lens at the polariser end. Landolt, in using the Mitscherlich instrument, determined the point of extinction by setting the dark band of complete polarisation central in the field, and this was the forerunner of the accurate method used by Landolt in the construction of his polarimeter (cf. *infra*). The mean error in the readings with the Mitscherlich instrument was about 0·1°.

The Soleil[2] double quartz plate, or "biquartz," was added to the polariser by Robiquet,[3] and the polarimeter was thereby made much more sensitive.

§ (2) THE BIQUARTZ.—Biot had observed that the yellow rays of white light were rotated through an angle of 90° by a plate of quartz 3·75 mm. thick. Soleil's biplate, which utilised this effect, consisted of two semicircular discs of quartz of that thickness, the one half being right-handed quartz (*i.e.* quartz that orientated the plane polarised beam in a clockwise direction as viewed by the observer) and the other left-handed quartz. The principle underlying its action can be seen from *Fig.* 2. Let AA′ be the

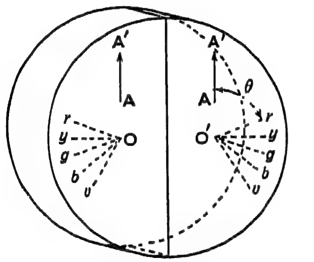


FIG. 2.

plane of polarisation of the white light entering the biplate. On emerging from the biplate the planes of polarisation of the red, yellow, green, etc., rays are indicated by dotted lines drawn from O and O′, the points of emergence of the rays. Owing to the structural difference in the two kinds of quartz, the planes of polarisation are rotated in opposite directions, the one phenomenon being a mirror image of the other. When the analysing nicol is parallel to the polariser, the yellow rays (which have been rotated an angle of 90° to AA′) will be cut out entirely. A portion of the light intensity of the red, green, and other rays will, however, be transmitted, the intensity transmitted being proportional to $\cos^2 \theta$, where θ is the angle that the plane of polarisation (of any particular wave-length), after emerging from the quartz, makes with that of the analyser. The resulting colour of both halves of the field is rose-lilac or amethyst.

In this position θ is, for the red rays, a little less than 90°, and for the green and green-blue a little more than 90°, while for all these rays $90 - \theta$ is not more than 45°.

Thus in the right half of the field a slight rotation of the analyser in a clockwise direction will cause the intensity of the red rays rapidly to increase, and that of the green and green-blue to diminish appreciably, the result being a very rapid change to a distinctly red hue. For a like reason the rotation causes the left half of the field to change as rapidly to a distinct blue.

Thus in the sensitive region where measurements are made the two halves of the field will have the same colour or tint only when the analyser is parallel to the polariser. If now an optically active substance be interposed, equality of tint can only be obtained by a rotation of the analyser to compensate for the rotation of the active substance. When the thickness of quartz is, as described above, so chosen that the yellow rays are extinguished by an analyser that is parallel to the polariser, the residual colour is exceedingly sensitive, turning either red or blue with a very slight rotation of the analyser. For this reason it has been called the sensitive tint, *teinte de passage*, or *Übergangsfarbe*. In order to have a complete explanation of why the 3·75 mm. quartz plate gives the most accurate results, the following factors must be considered: (1) Rotatory dispersion of quartz; (2) energy distribution of the white light spectrum; (3) relative sensibility of the average eye to radiation of different wavelengths. These together with the Malus $\cos^2 \theta$ law and the three primary sensation curves (red, green, and blue), such as given by Exner,[4] will give the change of tint with a given rotation of the analyser, by the method indicated by Ives.[5]

The disadvantages of this type are at once apparent: it must be used with white light, and the rotation obtained is approximately that for mean yellow light. Even if colourless substances are examined, there is no position of the analyser for which the tints in both halves of the field are the same, because of the rotatory dispersion of the substance itself, and thus systematic errors are introduced, of a kind extremely difficult to determine. This difficulty is considerably

[1] Mitscherlich, *Lehrbuch der Chemie* (4th ed.), 1844, cxi. 36.
[2] Soleil, *Compt. Rend.*, 1845, xx. 1805.
[3] Robiquet; cf. Sidersky, *Polarisation et Saccharimetrie* (2nd ed.), p. 48; Landolt, *Das optische Drehungsvermögen*, 1898, p. 294.
[4] Exner, *Sitz. Wien. Akad.*, 1902, abt. IIa, cxi. 857.
[5] Ives, *Journ. Frank. Inst.*, 1915, clxxx. 673.

increased when the substance is coloured. While it is almost impossible for a colour-blind person to use this method, there will be perceptible differences in the readings of normal persons, as the relative sensibility curves of individual eyes (to radiation of different wave-lengths) show considerable variation.

For these reasons the Soleil double plate is never employed in modern measurements.

§ (3) Jellett's Prism.—The next and probably the most important development in polarimetry is due to Jellett,[1] who utilised the photometric principle of matching two illuminated fields by varying their relative intensity. The eye is able, under favourable conditions, to detect very small differences of intensity (when no colour difference exists), and it is this principle that underlies the construction of all modern polarimeters and saccharimeters. A rhomb of Iceland spar is cut so as to form a right prism, as in *Fig.* 3. It is then cut in two in a plane B′BC parallel to the long edges of the prism and making a small angle a with the longer diagonal DD′ of the upper face. The one part is then

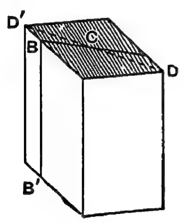


Fig. 3.

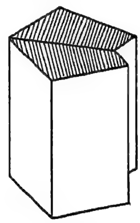


Fig. 4.

reversed, and cemented to the other as shown in *Fig.* 4. This composite prism is then mounted as the analysing prism, and the dividing line observed by means of a lens. This analyser receives a parallel circular beam from the polariser, and this beam is divided into four parts as in *Fig.* 5. The ordinary beams (one through each prism) remain central and symmetrical about the vertical interface; while the extraordinary images are refracted to the positions E and E′. These extraordinary images can be screened off by means of a suitable diaphragm at the eye side of the analyser.

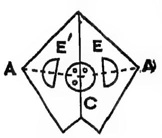


Fig. 5.

Let (in *Fig.* 6) CD represent the plane of polarisation of the polariser, and OA, OB the planes of polarisation in the two parts of the Jellett analyser. The two halves of the field will appear equally bright if the components of OA and OB on CD are equal, *i.e.* if OA is taken to equal OB, then they both make equal angles with CD. If now the analyser is rotated so that the polariser is in a position C′D′ with respect to OA and OB, the components of the latter (C′O, OD′ respectively) are no longer equal, the ratio of brightness being given, according to the law of Malus, as $(C'O)/(OD')^2$. It will be seen that under these conditions an equality of brightness in two halves of the field affords a sensitive means of determining optical rotation.

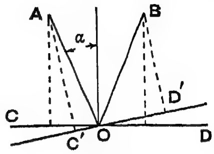


Fig. 6.

§ (4) Cornu's Prism.—Cornu[2] improved on Jellett's original prism, using a Nicol prism instead of the natural rhomb, the Nicol prism being divided in two in the same way as in Jellett's rhomb (*Fig.* 3). Instead of reversing the one half so as to have a resulting prism of the inconvenient and wasteful shape shown in *Fig.* 4, Cornu removed a thin wedge of spar, the apex of which was parallel to the long edges of the prism (see *Fig.* 7). The two parts were then cemented together without reversal, the planes of polarisation of the two parts making an angle with each other equal to the angle of the wedge of spar removed. This composite prism was placed at the polariser end, and a Nicol prism being used as analyser, the net result is much the same as that of the Jellett arrangement previously described, with the main differences that the double images of the Jellett prism are entirely eliminated, and that while Jellett used his prism as analyser, Cornu used his as polariser. This form of polariser is generally known as the Cornu-Jellett prism, and it remains in common use up to the present time for inexpensive polarimeters and saccharimeters.

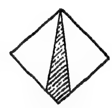


Fig. 7.

It is evident that the angle between the planes of polarisation of the Cornu-Jellett prism is fixed. This angle is called the "half shadow" angle, since to it is due that at the critical position of matching the field is neither dark nor yet completely illuminated. If this angle has been made comparatively small (say 2°-3°) in order to obtain greater sensitiveness, the prism may become no better than an ordinary nicol for examining the rotation of a light-absorbing substance, since the brightness of the field at the matching point, compared with the maximum brightness obtainable, is proportional to $\sin^2(a/2)$, where a is the half-

[1] Jellett, *Proc. Roy. Irish Acad.*, June 25, 1860; *ibid.* Jan. 22, 1863. The description of the use of this prism usually given seems to originate in an incorrect reading of *Rep. Brit. Assoc.*, 1860, ii. 13, and is incorrect.

[2] Cornu, *Bull. Soc. Chim. Paris*, 1870, xiv. 140.

shadow angle. Further, as will be seen later, it is not desirable to make this half-shadow angle too great, as the sensitiveness is reduced proportionately.

§ (5) THE LAURENT POLARIMETER.—In order to overcome this defect in the Cornu-Jellett polariser, Laurent [1] used a quartz plate covering half of a plane polarised beam from a Nicol prism. The principle can be understood from *Fig.* 8. The quartz plate A, whose optic axis lies in the plane of the paper, and has a direction OB, covers half of the circular aperture of the field of the Nicol prism in front of it. The thickness of the plate is chosen so that the difference between the paths of the ordinary and extraordinary rays, into which the incident plane polarised beam is divided, is some odd multiple of half the wave-length of the light that is used. If n_o and n_e are the refractive indices of the ordinary and extraordinary rays in the quartz for a given wave-length λ, then

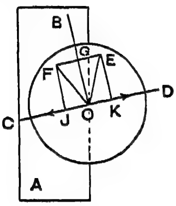


FIG. 8.

$$\delta = d(n_e - n_o) = (2m+1)\frac{\lambda}{2},$$

when d is the thickness of the plate, and δ the path difference introduced. For the sodium light D_1 and D_2 the minimum thickness is 0·0324 mm., but, as this plate would be too thin and fragile in practical use, some *odd* multiple of this thickness is chosen so that the plate is approximately ·5 mm. thick.

Let OE represent the plane of polarisation, and its length the amplitude of the linearly polarised homogeneous light coming from the polariser, whose plane of polarisation makes an angle $\theta/2$ with the optic axis OB of the quartz plate. The plane polarised beam, the direction and amplitude of which is given by OE, will, on entering the quartz plate, be divided into two parts, the one polarised in the plane of the optic axis OB and the other perpendicular to it in the plane CD. Thus OG and OK will represent the amplitudes and directions of the components at the moment that OE enters the plate. These components are in phase with each other.

Since the component of the beam represented by OK is accelerated in its passage through the quartz by an amount $(2m+1)\lambda/2$, m being an integer, as compared with the extraordinary ray OG, the position at emergence can be represented graphically by reversing the direction of OK, the components in phase on emergence being OG and OJ. Thus the resultant OF will represent the plane of polarisation after transmission of the beam through the quartz. If CD represents the plane of polarisation of the analyser, then a matching of the fields is obtained when the component OJ on CD is equal to the component OK from the unobstructed half of the field.

In order to vary the half-shadow angle, all that is required is to rotate the plane of polarisation of the polarising nicol; the angle that this plane makes with the optic axis of the plate is one-half the half-shadow angle.

This Laurent polariser, while allowing a variable half-shadow angle, is limited as to light source, the only permissible one being that for which the half-wave plate has been cut. The sensitiveness, as in all half-shadow polarisers, depends upon the magnitude of this angle, and on the intensity of the light source used. Heele [2] somewhat increased the sensitiveness by arranging the half-wave plate as a circular one mounted on a larger glass plate, as in *Fig.* 9 (i.), and Pellin has introduced the modification shown in *Fig.* 9 (ii.), when the retarding plate forms a ring around the centre of the field as shown.

FIG. 9.

When using a bright sodium light with a small half-shadow angle (about $1\frac{1}{2}°$) and taking the precaution suggested by Laurent of using a thin plate of potassium bichromate to cut off the other radiations of sodium, consistent settings can be made to within $\pm\frac{1}{2}'$.

Lippich [3] has, however, shown that the Laurent polariser cannot be relied on as an exact means of measurement. Since the relative importance of this type of polariser has become considerably smaller during the last twenty years, it will be sufficient to mention some of the main results of Lippich's analysis. The technical difficulties of making the plates perfectly parallel, with the optic axis lying exactly in the same plane, and of exactly the correct thickness for the particular wave-length chosen, are with modern methods perhaps not insuperable. But Lippich found that they certainly were not made with sufficient accuracy; and even had they been ideally perfect it would have been necessary, in order that different instruments might agree, that the thicknesses of all Laurent plates should be identical, and not merely *any* odd number of half-wave-length retardation. Again, even if agreement were thus secured, the precise significance of rotation measurements depended on a knowledge of the rotatory dispersion of the substance under test, which knowledge could not be obtained on instruments of the Laurent type.

Finally, one and the same instruments give different measurements of a rotation with different half-

[1] Laurent, *Journ. d. Phys.*, 1874 (I.), iii. 183; 1879, viii. 164; Dingler's *Poly. Journ.*, 1877, ccxxiii. 608.

[2] Heele, *Zeitsch. Instkde.*, 1896, xvi. 260.

[3] Lippich, *Sitz. Wien. Akad.*, 1890, abt. IIa, xcix. 695.

shadow angles. All these faults result from the imperfect homogeneity of all practical light sources.

Schulz[1] has recently examined the errors due to elliptic polarisation caused by the obliquity of the rays falling on the Laurent plate. He finds that if A is the diameter of the aperture or diaphragm at the polariser and B that of the analyser, and L is their distance apart, then if $A+B=\cdot023L$, the error caused by obliquity of the rays is not greater than $\cdot001°$. If A and B are to be the same, then for a polarimeter to take 200 mm. tubes (L=230 mm.) the value of $A=2\cdot6$ mm., for a 400 mm. instrument (L=430 mm.) A must not exceed 4·9 mm.

Notwithstanding the above criticisms, it cannot be doubted that the large discrepancies usually found in results with different Laurent polarimeters must be mainly due to defective manufacture and lack of purity of the quartz.

§ (6) Lippich's Polariser.—The polarising system of Lippich,[2] which is in general use in the best polarimeters at present, combines the advantages of a variable half-shadow angle, and freedom to use either a heterogeneous light source or homogeneous light of any desired wave-length.

In the Lippich system an ordinary nicol A (*Fig.* 10) is followed by a smaller nicol B covering half the field. The planes of polarisation of the prisms A and B make an angle θ with each other, and thus it is possible to employ the photometric principle of Jellett in measuring rotations, as has been previously described. In order to have a field of uniform brightness at the matching point, the nicol B must be so designed that all the rays followed back from any point within the diaphragm MN of the analyser must pass into A without any partial obstruction by the small prism B.'

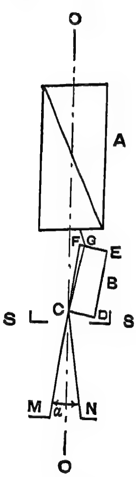


Fig. 10.

If $\alpha/2$ is the half-cone angle subtended by the analyser diaphragm at C, then in order that no light from A shall enter the analyser after external reflection at the surface CF the latter must make an angle with the axis OO′ slightly greater than $\alpha/2$ (say $(\alpha/2)+\beta$). If the ray NC traced back from the analyser is to pass through B, without internal reflection, then in order that the ray CG should make a small angle ϵ with the side CF the angle FCD must be suitably chosen. It is found that if ϵ and β are made about 10′, the angle FCD must be between 92° and 94°. This small prism is commonly called the "half Lippich prism."

Lippich improved on his original two-field system by having two half Lippich prisms B and C mounted as shown in *Fig.* 11. The resulting field (DEF) is divided into three parts.

The original intention of Lippich was that the planes of polarisation of the half prisms B and C should be slightly inclined to one another, so that if a match is made between D and E, the difference in brightness between E and F is just discernible. By this means the range of uncertainty in setting should theoretically be halved.

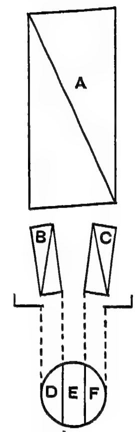


Fig. 11.

Perhaps the greatest advantage of the three-field system, however, lies in the fact that it enables the observer to correct any slight error in the alignment of the light source with the optical axis of the polarimeter. The distribution of the intensity of illumination over the field is liable to change rapidly with a slight displacement of the light source, and an error due to this cause would generally be far greater than that due to the lack of sensitiveness of the eye, especially if the observer has had some practice in making polarimetric measurements.

Lummer[3] has constructed a polariser with a field divided into four equal parts. This is done by adding to a double-field Lippich system two more very small half Lippich prisms each of which covers a field equal to a half of prism B in *Fig.* 10. These are placed at the outer edges so that the field is divided into four equal parts. In this polariser it is not a condition of uniform brightness that is aimed at, but the equally distinct appearance of two of the parts on a uniform background.

Whether this type of polariser shows any real gain of sensitiveness over the Lippich three-field polariser does not appear to have been adequately tested. The arrangement is too complicated ever to come into general use, unless its advantages were very real. In practice it is found difficult enough to adjust, and keep in adjustment, the Lippich three-field type.

§ (7) Variation of Sensitiveness with the Half-shadow Angle.—It has been mentioned that the sensitiveness or accuracy of setting with the photometric matching principle of Jellett depends on the half-shadow angle. It is generally assumed that the eye can recognise a difference of brightness of about 1 per cent when examining two adjacent fields.

[1] Schulz, *Phys. Zeitschr.*, 1920, xxi. 33.

[2] Lippich, *Lotos*, 1880, ii.; *Zeitschr. Instkde.*, 1882, ii. 167; 1892, xii. 333; 1894, xiv. 326; *Wien. Ber.*, 1882, lxxxv. 268; 1885, xci. 1059; 1890, xcix. 695; 1896, cv. 317.

[3] Lummer, *Zeitschr. Instkde.*, 1896, xvi. 209.

Let OA and OB (in *Fig.* 12) represent the planes of polarisation in each part, and their lengths represent their amplitudes which are assumed equal, and let AOA represent the plane of polarisation of the analyser when the two fields are equally bright. Then, if the analyser be rotated a very small angle δ into the position A′OA such that the beam originally polarised in the direction OA is, after passage through the analyser, just perceptibly brighter than that of the other half of the field, OA and OB will then make angles of $90-\theta-\delta$ and $90-\theta+\delta$ with the new plane of polarisation of the analyser. The ratio of the intensities in the two halves will be, according to the law of Malus:

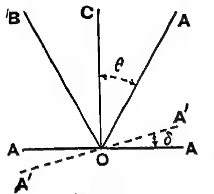


FIG. 12.

$$\frac{\sin^2(\theta-\delta)}{\sin^2(\theta+\delta)}=p,$$

which gives

$$\delta=\frac{(1-\sqrt{p})}{(1+\sqrt{p})}\tan\theta.$$

(In obtaining this expression, which gives δ in circular measure, δ is taken to be so small that $\cos\delta=1$ and $\sin\delta=\delta$.)

If p is taken to be ·99, which is equivalent to saying that 1 per cent change of intensity can be detected, then

$$\delta=\cdot 0025\tan\theta.$$

Remembering that θ in *Fig.* 12 is half of the half-shadow angle, the solid line in *Fig.* 13

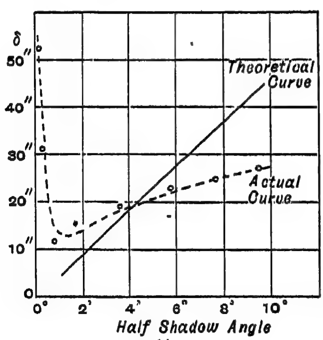


FIG. 13.

represents the kind of variation in sensitiveness with half-shadow angle that is to be expected on theoretical grounds. The actual curve of mean error (dotted line) given by Schulz and Gleichen [1] indicates that with a half-shadow angle of about 9°-10° the accuracy of setting is considerably greater than that shown by the theoretical curve. This shows that under favourable conditions the eye can detect a change of intensity of illumination of about ·75 per cent.

It is usually assumed that the cause of the rapid loss of sensitiveness as the half-shadow angle becomes less than 1° is that the general illumination then becomes too faint for accurate matching. The chief reason, however, is that with such half-shadow angles the effect of the elliptic polarisation becomes appreciable. The light incident on the analyser is assumed to be completely *plane* polarised, whilst in practice it is always to a slight extent *elliptically* polarised, to an extent depending on the total amount, and azimuth of the double refraction, present, between the oil or balsam film of the polariser and that of the analyser. There is also always a slight admixture of depolarised light. The effect as seen through the analyser can be regarded as merely adding an equal amount of general illumination to each field. This is not appreciably altered as the half-shadow angle is made to approach 0°, and as the intensity of illumination from the plane-polarised component becomes a smaller and smaller proportion of the total intensity of the field, the analyser has to be rotated a correspondingly greater amount before the eye can discern a change of intensity. This can be easily verified by observing the sensitiveness of a polarimeter with a half shadow of 2°-3° using a very poor light source, so that the field at the matching point is very dark. With such an arrangement a much higher accuracy will be attained than with a half-shadow angle of, say, ½°, even when in the latter case the light source is so improved that the intensity of illumination at the matching point is considerably brighter than with the larger half-shadow angle.

It can be shown that in consequence of the inequality of the intensity of illumination in the two halves of an ordinary Lippich polariser, a systematic error is made in the measurement of a rotation either if the substance under examination is slightly bi-refringent or if a small amount of accidental double refraction is otherwise introduced into the path after the zero reading of the polarimeter has been taken.

This has not yet been fully investigated, but it is found that with a half-shadow angle of about 3° and an ellipticity of 1/100 (ratio of minor to major axis) the error is of the order of 1′.

§ (8) WILDE'S POLARIMETER.—Other forms of polariser have been suggested from time to time, none of which, however, have come into general use.

In the form of polarimeter devised by

[1] Schulz und Gleichen, *Polarisationapparate und ihre Verwendung*, Stuttgart, 1919, p. 66.

Wilde,[1] both the polariser and analyser are ordinary Nicol prisms. In order to obtain an accurate setting, a Savart plate is mounted in front of the analysing nicol. As this consists of plates of Iceland spar or quartz, the optic axes of which make an angle of 45° with the direction of the light, and whose principal sections are at right angles, a system of interference fringes is seen in general. When the plane of polarisation of the beam from the polarising nicol coincides with the principal section of either of the plates, the fringes disappear at the centre of the field. By reducing the intensity of the light source, the breadth of the band showing no fringes can be considerably narrowed, and this can then be set symmetrically with respect to a pair of cross lines placed at the focal plane of two lenses that form a telescopic system interposed between the Savart plate and the analysing nicol.

The instrument can be used with a homogeneous light source of any wave-length, and has the great advantage that a strong light source is not required. Considerable practice is required before the observer can obtain an accuracy of setting comparable with that of a half-shadow polariser, as the termination of the fringe system is not sharply defined. For this reason the Wilde instrument is very rarely used.

§ (9) Landolt and Lippich.—Landolt[2] observed that when a very intense light source is examined by a pair of wide-angled Nicol prisms, the field, instead of being uniformly dark, is crossed by a narrow black band only. This was first explained by Lippich,[3] who showed that, due to the varying obliquity of the rays incident on the prism, the directions of vibration at different parts of the field were not strictly parallel. He showed further that in the case of Nicol prisms whose end faces were perpendicular to the axis of the prism the direction of the vibration was represented by a system of converging lines which met at a point P outside the prisms. The direction of the plane of polarisation will therefore be given by arcs of circle with P as centre. In *Fig.* 14 the solid arcs *abcd*, etc., will indicate the direction of the plane of polarisation at any point in the field.

A similar state of affairs holds in the case of the analyser, when the *dotted* arcs *a'b'*, etc., will represent the direction of the planes of polarisation of the analyser at various points in the field.

It will be noticed that these two sets of curves are only perpendicular to each other along the locus FG-J, and therefore complete

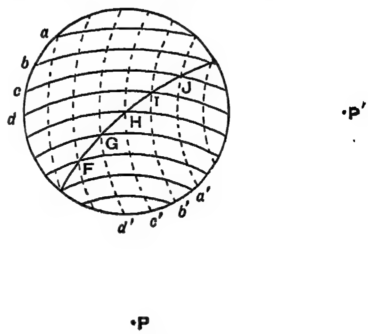


Fig. 14.

darkness will occur only along this line. A slight rotation of either the analyser or polariser will obviously change the position of this locus. Lippich made use of this fringe in the construction of a polarimeter, when the position of the fringe with reference to a pair of cross-hairs served as the criterion instead of the matching of fields as in a half-shadow polariser. By this means Lippich states that it is possible to set the analysing nicol to within two or three seconds of arc,[4] an accuracy considerably greater than that obtained by any type of half-shadow polariser. In order to see the fringe, the light source must be very intense (*e.g.* sunlight or the electric arc), and so the method cannot be conveniently applied for polarimetric work, since it is difficult to obtain the necessarily homogeneous light source of the required intensity.

§ (10) Lummer's Polarimeter.—In the polariser evolved by Lummer,[5] plane polarised light falls on a 45° prism the hypotenuse face of which has silvered strips S (*Fig.* 15). The light enters the prism normal to the face ABCD. The portion falling on the unsilvered part of the hypotenuse face ABEF is totally reflected, while the light falling on the silvered surface suffers metallic reflection. If the plane of polarisation of the incident beam is parallel or perpendicular to the plane of reflection, the planes of polarisation of the reflected beams (total reflection and metallic reflection) will be coincident. If the polarising nicol is rotated through an angle θ, the portions of the beam suffering total reflection will be rotated an angle $\theta°$ from its original azimuth,

[1] Wilde, *Über ein neues Polaristrobometer*, Bern, 1862; *Vierteljahrsschrift der Naturf. Ges.*, Zürich, 1898, xliii. 57; 1899, xliv. 136.

[2] Landolt, *Das optische Drehungsvermögen organischer Substanzen* (1st ed. 1879), p. 95.

[3] Lippich, *Sitz. Wien. Akad.*, 1882, lxxxv. 268. Later investigation has shown that the analysis of Lippich is only a close approximation. Cf. Borek, *Verh. d. D. Phys. Ges.*, 1919, xxi. 338.

[4] G. Bruhat and M. Hanot claim that it is not possible to measure a rotation of the order of 20' by this method with an error less than 1'. With intense sources, *e.g.* mercury arc, this error can be reduced by half. *Acad. Science*, Paris, May 30, 1921.

[5] Lummer, *Verh. d. Ges. Deutsch. Naturf. u. Ärzte Wien*, 1894 (II.), i. 79; *Zeitsch. für Instkde.*, 1895, xv 293.

while that reflected from the silver strips will be rotated through an equal angle in the opposite direction. A half-shadow arrange-

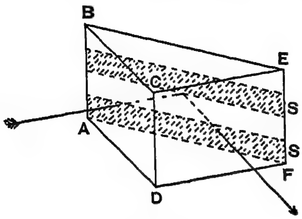


FIG. 15.

ment is thus obtained, as in the case of the Lippich system, with a half-shadow angle of 2θ, which can be varied as required.

Owing to the comparatively large amount of glass in the path, the effect of small amounts of double refraction, which is negligible in thin plates, now becomes appreciable and the field at the matching point is not uniform, having brighter or darker patches in the field of view. With a large half-shadow angle, the linearly polarised light becomes elliptically polarised, and, as can be shown from the electromagnetic theory, the ellipticity will be different in the two cases. For half-shadow angles as used in practical work, the ellipticity introduced at reflection is negligible, the only drawback being the double refraction due to lack of homogeneity and annealing of the glass.

It is possible that, as the technique of optical glass manufacture is improved, this type of polariser may come into common use. A detailed investigation of its properties has been made by Volke.[1]

§ (11) GENERAL CONSIDERATIONS.—A large number of polarisers have been devised which depend on the double refraction or rotatory power of quartz for their half-shadow effect, and a typical example, the Laurent plate, has been described in detail.

A most serious objection to this class of polariser is the great scarcity of optically good quartz, the purity of which immediately sets the limit to the accuracy and sensitiveness of the particular polariser. From this cause alone a considerable variation is to be expected between the performances of individual polarisers of any one type.

(i.) *Poynting's Polariser.*—The polariser of Poynting[2] is made by dividing a circular plate of quartz, cut perpendicular to the optic axis, along the diameter. One half is slightly reduced in thickness and the two halves reunited. The half-shadow effect is obtained by the difference of the rotation produced in each half of the plate. The half-shadow angle is therefore a fixed one, and furthermore it is necessary to employ a homogeneous light source to avoid the effects of rotatory dispersion.

(ii.) *Nakamura's Polariser.*—Nakamura[3] has shown that maximum sensitiveness of a Soleil double quartz plate is obtained when the plate is approximately ·4 mm. thick, instead of 3·5 mm. or 7 mm. as originally used. Wright[4] independently arrived at the same conclusion, and as this polariser admits of only one fixed half-shadow angle, the latter[5] has evolved an ingenious application, which not only affords a variable half-shadow angle, but also gives this variation without a change of the zero position of the analyser—a most important consideration in the case of saccharimeters. It consists of two wedges of quartz of equal angle but of opposite rotation, the one mounted above the other as shown in *Fig.* 16, and immediately behind each half

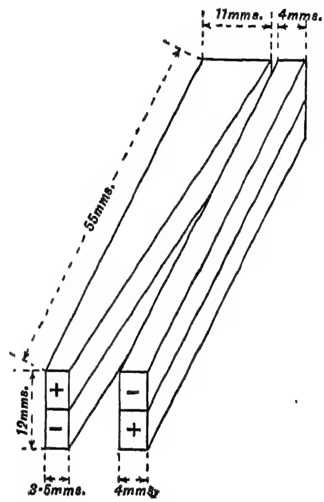


FIG. 16.

wedge is a parallel plate of quartz of opposite rotation. The dimensions suggested by Wright for polarimetric work are given in the figure. It will be seen that at a point 3·3 mm. from the thinner end of the wedge the total rotation is zero in both parts, since the thickness of the double wedge is then the same as that of the double plate of opposite rotation. The field of view when the nicols are crossed will therefore be comparatively bright, except for a narrow dark band at this point, which will run vertically across the field. If either nicol is given a slight rotation the part of the dark

[1] Volke, *Inaugural-Dissertation*, Breslau, 1909. See also Volke, *Ann. d. Physik*, 1910, xxxi. 609.
[2] Poynting, *Phil. Mag.*, 1880 (5), x. 18.

[3] Nakamura, *Centralblatt für Min.*, 1905, pp. 267-279.
[4] Wright, *Amer. Journ. Sci.*, 1908, xxvi. 377.
[5] *Ibid.* pp. 391-398.

band in the upper wedge will move rapidly in one direction, whilst the dark band in the lower wedge will move a corresponding distance in the opposite direction. In using this device an approximate setting is first made, and biquartz wedge is then placed in the path, the accurate reading being obtained by rotating the analysing nicol until the black bands are brought in alignment.

As the thicker end of the double wedge is moved into the field, the coincidence of the two parts of the black band remains unaltered, the field itself becoming generally brighter while the band itself is not so sharply defined. With wedges of the dimensions given in *Fig.* 16, illumination, equivalent to that given by a Lippich polariser of 15° half-shadow angle, is obtained when the thicker end of the wedge is in the field of view. As only one magnifying lens is needed, the usual light-absorbing observing telescope can be dispensed with. This form of polariser has not yet been employed in any commercial polarimeter or saccharimeter.

(iii.) *Brace's Polariser.*—Probably the most sensitive and accurate form of half-shadow polariser is that of Brace.[1] Essentially it consists of two plates of mica mounted between crossed nicols. The first consists of a strip (about ·00017 mm. thick or $\frac{1}{100}\lambda$ retardation) covering part of the field, and is fixed; the second "compensator" plate covers the whole field and has a retardation of $\frac{1}{4}\lambda$. Two positions of the compensator plate can be found at which the intensity of the light is uniform over the whole field. By varying the angle which the principal section of the strip makes with the azimuth of the polariser (or analyser) a variable half-shadow effect is obtained. For a complete account of the principles underlying its use, reference must be made to the original papers. The main advantages of this polariser are that the separating line is vanishingly fine, and that the double refraction due to strain between the polariser and analyser is reduced to a minimum, so that it can be successfully used with what is equivalent to a very small half-shadow angle. It is mostly used for determination of ellipticities; with a very thin sensitive strip a change of phase of $6{\cdot}2\times10^{-7}$ can be detected under favourable conditions. Bates[2] has applied this type of polariser to a spectra polarimeter and found the extreme differences of readings less than ·007°. The Brace system is, however, rather fragile, and its use is not recommended except in work requiring the highest obtainable precision.[3]

§ (12) CONSTRUCTION AND MECHANICAL DESIGN OF POLARIMETERS. *Simple Types.*—Since the requirements in polarimetry are so varied, it is only to be expected that the various types available should be numerous. For many purposes, such as urine analysis, an accuracy of measurement of ·1° is sufficient, and a simple instrument such as is shown in *Fig.* 17 is all that is required. The polarising system P is connected to a divided circle C

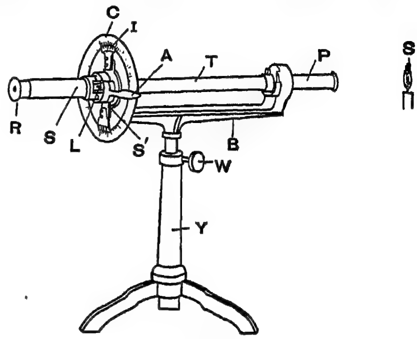


FIG. 17.

by means of the bracket B. The analysing system can be rotated by means of the lever A, and its azimuth is given by the index arm I. The observation tube T rests on suitable supports at each end, and the whole instrument is supported by a tripod stand Y.

In making an observation, the sodium light source S is either focussed on the polariser end P or brought as near to it as is possible without unduly warming the polariser, the tube T is filled with water and placed in position, the eyepiece R of the observing telescope is either pushed in or drawn out, until a sharp image of the separating line of the polariser is seen. The arm A is rotated until a position is reached when the two fields are *equally dark*. If the index arm I does not then read exactly 0° or 180°, the procedure is as follows. The analysing nicol is mounted in a tube with a short arm L fastened to it. This tube again in turn fits in the outer tube to which A is fixed, and the analyser can be rotated a small amount with respect to it by means of thumb-screws which bear on L in the manner shown. The arm A is moved, therefore, until the index arm reads 0, and the thumb-screws turned until a match is obtained. The tube T is now filled with the liquid to be examined, and in general both parts of the field will appear bright. The arm A is again rotated until a position of equal darkness is obtained. The new reading of the index arm will give the rotation of the substance.

In some forms the whole instrument can be inclined, and clamped in any position by means of a swivel joint and bolt. This obviates the necessity of raising or lowering the light source to suit the instrument. The circle C is

[1] Brace, *Phil. Mag.*, 1903, v. 161; *Phys. Rev.*, 1904, xviii. 70; *Phys. Rev.*, 1904, xix. 218.
[2] Bates, *Ann. d. Physik*, 1903 (4), xii. 1096.
[3] *Circl. Bur. Sta.*, 1918, No. 44, p. 12.

generally about 3 in. in diameter and divided into degrees; the possible error in measurement is therefore ·1° or ·2°. When slightly greater accuracy is required, a vernier reading to 0·1° is engraved on the arm I instead of the single fiduciary line, and it is often the practice to mount a magnifying lens on this movable index arm, with a small mirror attached to it that projects outside the circle and reflects on to the scale and vernier enough light from the source *s* to make a reading possible in an otherwise darkened room.

The optical arrangement is shown in *Fig.* 18. The source *s* being either focussed on or close to the aperture A, the latter can be considered as the virtual light source. The position and focal length of the condenser C are chosen so that the image of the stop A falls on the objective M of the observing telescope. The polarising system immediately follows the condenser. In the figure the Laurent polariser is shown, D being the polarising nicol, and E the half-wave plate covering a part of the field. In this type of instrument having a fixed half-shadow angle, the Cornu-Jellett polariser is often employed instead of the Laurent. A second diaphragm F limits the beam and prevents stray illumination by reflection from the sides of the observation tube GH. The calcite prisms are difficult to clean without scratching; a thin protecting window K′ is therefore mounted in the aperture F, and a similar one K at J. A low-power telescope MN immediately follows the analyser L, and this is adjusted to give a clear image of the separating line of E in the field of view.

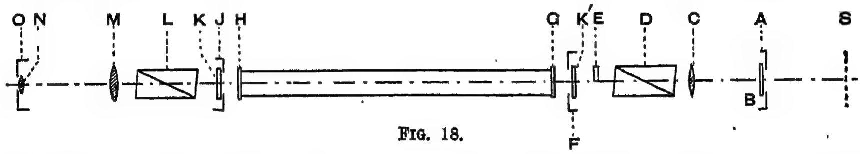


FIG. 18.

Since the radiation from a sodium burner consists of lines in blue and other parts of the spectrum as well as the strong yellow radiation, a filter B (*Fig.* 18) of potassium bichromate is placed close to the aperture A. This usually consists of either a thin plate (about 1 mm.) of potassium bichromate crystal balsamed in between two glass cover plates, or a glass cell containing a solution of the bichromate in water.

The instrument described above is as manufactured by Pellin; in a corresponding simple polarimeter of Schmidt and Haensch, a Lippich polariser with a fixed half-shadow angle is employed.

In the corresponding polarimeter of Bellingham and Stanley, which reads by vernier to ·05°, the polariser consists either of a Laurent plate or a modified Jellett prism. In the latter case, instead of reversing the one half prism as described by Jellett (see § (3)), two small natural rhombs are taken, and a section *abcd* (*Fig.* 19) is cut with the long sides *ad* and *bc* parallel to the cleavage edges of the rhomb, but with *ab* and *cd* making an angle $90° - \theta$ with the principal section EF which contains the optic axis of the crystal; while in the second rhomb *ab* is made to give an angle $90 + \theta$ with EF. When these are joined or cemented together with their corresponding faces *abcd* in contact, the principal section of the one half makes an angle of 2θ with that of the other. In order to have a sharp separating line, the one rhomb is placed a few millimetres behind the other so that the single edge *ab* of the one rhomb forms the separating line. The corner G of each rhomb is ground into a flat surface parallel to the line *cd*, but making a small angle with the plane *abcd*. This allows greater latitude in adjustment without the extraordinary ray from G coming into the field of view. This polariser forms a single detachable unit that can be quickly interchanged.

The circle is bevelled at 45° to its own plane, and the vernier is engraved on another fixed circle similarly bevelled.

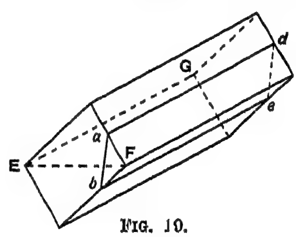


FIG. 19.

The graduated scale and vernier therefore lie in the same surface, and no parallax error can be introduced. Since this vernier is engraved on a second complete circle instead of on a narrow index arm, the danger of widening the space between the scale and vernier, caused by accidental knocks, is considerably lessened. In this instrument the vernier is fixed while the circle rotates with the analyser. In order to avoid the accidental tilting of the instrument, which is liable to occur in the type shown in *Fig.* 17 when the thumb-screw of the swivel *w* has not been sufficiently tightened, the girder connecting the polariser and analyser is furnished with two bored holes into either of which the upright of the tripod can be placed. The one

supports the instrument in a rigid horizontal position, and the other inclines the polariser end downwards, at an angle of about 15°. This instrument has a movable trough, consisting of two close-fitting circular tubes. Except for a narrow portion at each end, nearly a half of each tube is cut away, so that the tube holding the liquid to be examined can be inserted. When the inner tube is now rotated by a suitable knob through an angle of approximately 180°, the observation tube is completely enclosed. The chief advantage of this kind of trough is that no hinges are required; these often corrode after coming in contact with some of the liquids used in polarimetric analysis.

The vernier is engraved to read directly to ·05°.

§ (13) HIGH ACCURACY POLARIMETERS.—When an accuracy of ·01° is required, several changes are necessary, the most important being that of eliminating the eccentric error in the circle graduation. After the circle has been turned, it is usually mounted on the spindle of another machine for dividing, consequently the common centre of the radial graduation lines will not coincide with the true centre of the circle when a small particle of matter is lodged between the spindle and the sleeve of the circle. Unless exceptional care is exercised, the error in measuring an angle of about 90° may amount to ·02°. If two verniers are used, mounted approximately at 180° to each other, this error is eliminated if the *mean* value of a rotation (as given by each vernier) is taken as the true rotation. This will hold good for eccentricities up to any likely to occur in practice. When the use of the polarimeter is not confined to a special purpose, a variable half-shadow angle becomes almost a necessity. Furthermore, it is desirable to make the instrument as rigid as possible, and since a small change from alignment with the light source may cause an appreciable error in the measurement, the instrument should stand firmly on the table or bench. For these reasons it is usual to design the more accurate instruments to have trestle mounts with a substantial base. In addition, the space surrounding the optical axis of the instrument, between the polariser and analyser, should be as free as possible, to allow for the interposition of electromagnets, thermostats, or electric furnaces, as may be required.

In order to show the mechanical construction of a modern accurate polarimeter, the construction and setting of one instrument will be described in detail and the more important differences in the design of other manufacturers will be noted.

(i.) *Adam Hilger, Ltd.*—In the polarimeter illustrated in *Fig.* 20 (made by Adam Hilger, Ltd., London) the half-shadow angle of the polariser is varied by means of the small lever arm A which moves behind a small fixed scale giving directly the half-shadow angle; by means of the small thumbscrew shown, it can be clamped in any position. The verniers GG at the analyser end are fixed while the circle can be rotated with the analysing nicol. The verniers are observed by means of the low-power eyepieces MM.

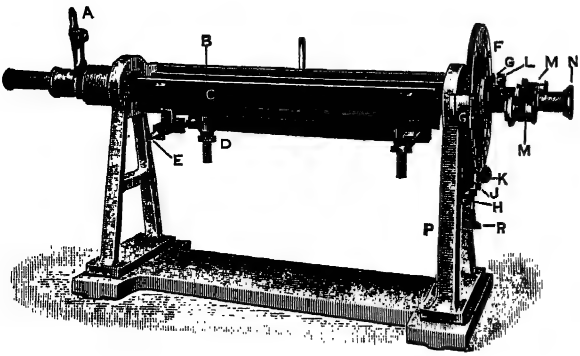


FIG. 20.

A monochromatic light source is placed in alignment with the optical axis of the instrument, and in such a position that a real image of it is formed approximately at the analyser. The eyepiece N of the low-power observing telescope is moved in or out until the edges of the half Lippich prisms at the polariser are sharply in focus. The circle F is set at the zero position. If the three parts of the field are not of equal intensity, the analyser is slightly rotated by means of the thumb-screws at L while the circle remains stationary.

The trough C can be raised or lowered or slightly tilted by means of knurled nuts on the screws DD. If a large observation tube is to be used the trough must be correspondingly lowered so that the centre of the tube corresponds with the optical axis of the instrument.

The observation tube being placed in position, the circle F will have to be rotated to a new position before a match of the fields is obtained. When this position has been found approximately the screw head R is tightened. A ring with a projecting arm J fits smoothly on a collar that forms an integral part of the circle. When R is tightened, this ring, and consequently the arm J, is clamped to the circle.

The latter can therefore be slowly rotated by turning the screw K which bears against the arm J. When the end of the screw K is moving away from J, the latter is kept in contact with the screw by pressure of the spring and plunger H on the other side. A slow motion arrangement such as the one explained above is practically a necessity when settings have to be made to within ·01°.

The trough C slides on the pins E for centring, and can be quickly removed when required. If necessary, the bar B connecting the polariser and analyser can also be unscrewed and removed, as the trestle mounts PP are sufficiently rigidly attached to the massive iron base S.

(ii.) *Bellingham and Stanley.*—In the corresponding Bellingham and Stanley (London) polarimeter special attention has been devoted to the illumination of the circle, as shown diagrammatically in *Fig.* 21. The scale and

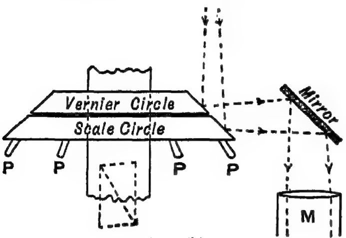


FIG. 21.

verniers are turned conically away from the observer, and are illuminated by light from the source. A plane mirror set approximately parallel to the face of the vernier and scale, as shown, serves to reflect the light into the low-power eyepiece M. A similar arrangement (not shown in the diagram) is mounted on the opposite side so that the eccentric error can be eliminated. The scale circle can be rotated and set approximately by means of the capstan heads PP, the final adjustment being made with a slow motion similar in principle to that already described.

For very accurate work it is necessary to have the room in complete darkness, so that the condition generally known as the "dark adapted eye" may be obtained. All the light from the source other than that entering the polariser must then be screened off, and the scale and verniers have to be illuminated independently as in other instruments. But it is an appreciable advantage of this arrangement that the graduations do not become clouded with moisture from the breath of the observer, as the scale is turned away from him. When a variable half-shadow polariser is required with this instrument, a modified Lippich system is supplied instead of the Jellett arrangement previously described. Here the long rhombs of spar (that form the components of the Jellett prism) are placed in the usual positions of the half Lippich prisms, and the large nicol is rotated to vary the half-shadow angle as usual. A circular trough is used instead of the V-shaped trough of the Hilger instrument; this can be quickly removed, and two adjustable V-shaped supports can be used to hold an observation tube that is either too large or too small for the trough.

(iii.) *Schmidt and Haensch.*—Corresponding polarimeters of Goerz and Schmidt and Haensch (Berlin) are generally similar in design to the Hilger instrument described above, except that the circles in these instruments are completely encased so as to keep the scales from tarnishing in a chemical laboratory. In the Goerz instrument the parts PP and S of *Fig.* 20 are one casting, PP being massive iron pillars.

It is also customary to fit a small circular table on the base S, which can be raised or lowered as required.

The large polarimeter made by Schmidt and Haensch at the instance of Landolt has some interesting features. The trestle holding the analyser circle is about two inches thick, so as to ensure a smooth-running circle; the bush in which the rotating tube from the circle turns is long enough to ensure that there will be no sagging of the circle due to the projecting weight. The optical axis of the instrument is only about 6″ above the upper face of the base, compared with 9″ as is usual, thus making the instrument exceptionally rigid. Instead of a single trough, a series of four are mounted side by side on a common axis, so that any one trough can be quickly brought into the optical axis of the instruments. This serves for the rapid intercomparison of the rotations in two or more observation tubes without the necessity of handling them and the consequent temporary rise in temperature.

§ (14) SPECTRO-POLARIMETERS.—It is often necessary to determine the rotation of substances for various wave-lengths that cannot be obtained singly with the same ease as the yellow doublet of sodium. In the method devised by Fizeau and Foucault,[1] a slit is placed at the polariser of a simple polarimeter (consisting of two Nicol prisms and a graduated circle) with a prism and telescope. Sunlight is used as the light source. The nicols are crossed in the absence of the optically active material; when the latter is placed in the path a spectrum is seen at the telescope. On rotating the analyser, a dark band will be seen to move across the spectrum, and the rotation of the substance for a given wave-length is given by the amount that the analyser has to be rotated until the black band is at the position of that wave-length in the spectrum.

[1] Fizeau et Foucault, *Compt. Rend.*, 1845, xxi. 1155.

The method has been improved by Lippich[1] and others, but it only gives good results when the rotatory dispersion is large.

A convenient method of accurately determining the rotatory power of substances for different wave-lengths is shown in *Fig.* 22, the arrangement being similar to that used by Lowry[2] in his investigation on the rotatory dispersion of quartz. The light source S is placed at the focal plane of the short focussed

plane of polarisation is parallel to that of the analyser, there is no appreciable reduction of intensity on transmission through the analyser.

It follows, therefore, that in order that these radiations may not influence the sensitiveness or the position of matching for the particular wave-length that is desired, the proportion passing through the slit must at most be of the order of ·1 per cent. Owing to slight scattering at the refracting surfaces of the

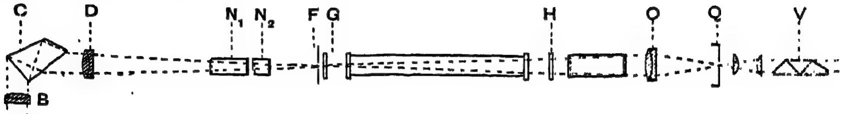
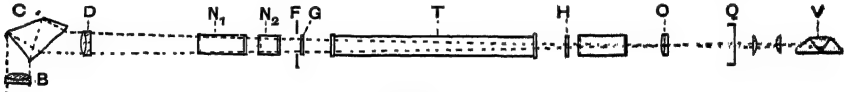


FIG. 22.

condenser W, which is close to the slit A, so that the latter is uniformly illuminated. To secure brightness of the final image the focal length of the condenser is so chosen that the light from S (which has a definite magnitude) fills the objective B of the collimator, or so much of it as is operative in forming the final image. The parallel beam emerging from B enters a constant deviation prism C, and the resulting spectrum is focussed at the slit F of the polariser by means of the long focussed objective D. The Lippich

deviating prism, lenses, etc., this spectral purity can never be attained in practice. It is therefore desirable to use a second spectroscope in the form of a direct vision prism V at the analyser. If, for example, a quartz mercury lamp be used as the source, and it is desired to observe the rotation of a substance for the green line (5461), this line is set on the slit F of the polariser. When the analyser is rotated to the matching position, and the slits F being narrow, the green band as seen through N will be accompanied by slightly fainter bands of yellow and violet. When a rotatory dispersive medium is placed between



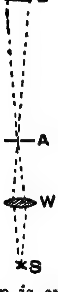


FIG. 23.

polariser N_1N_2, which is usually of the three-field type, is arranged so that the separating lines are horizontal. G and H are glass protecting covers for the polariser and analyser nicols.

When the fields are matched with a half-shadow angle of $1\frac{1}{2}°$, the intensity of illumination is only about ·1 per cent of the intensity when the polariser and analyser are parallel (theoretically it is ·06 per cent, but a slightly higher percentage is transmitted owing to the almost inevitable elliptic polarisation in the light when it reaches the analyser). When an active substance is placed in the tube, the planes of polarisation of other wave-lengths may lie in azimuths considerably different from that of the analyser, so that the latter only reduces their intensities slightly. In the extreme case of a particular wave-length whose

the polariser and analyser, these may become so important as to prevent an accurate reading of the rotation of the green ray, as described in general terms above.

One of the writers has found that somewhat better results are obtained with an arrangement as shown in *Fig.* 23. Here the light source S is *focussed on* the slit A by means of the condenser W, and the light of a particular wave-length emerging from the prism C is focussed by means of the lens D on the objective O of the observing telescope (the usual condition of illumination in polarimetry). The whole of the field at F is uniformly illuminated, and the slit can be opened until the yellow and green bands (supposing a mercury light source is used) as seen with the direct vision prism nearly overlap.

The diaphragm Q in *Figs.* 22 and 23 serves to eliminate the effects of internal reflection in the observation tube.

The mechanical construction of the spectropolarimeter, made by Adam Hilger, Ltd., can be seen from *Fig.* 24. The monochromator,

[1] Lippich, *Wien. Sitzungsber.*, 1882 (II.), lxxxv. 307.
[2] Lowry, *Phil. Trans.*, 1912, ccxii. 261.

polariser, and analyser form separate units that can be clamped in any position along the girder bed. In the illustration they are arranged for a trough to take observation tubes up to 100 cm. long. The drum of the monochromator has a long rod attached to it so that various wave-lengths can be set from the analyser end. While it is not absolutely necessary, it is best to use a direct vision prism that does not greatly deviate the par-

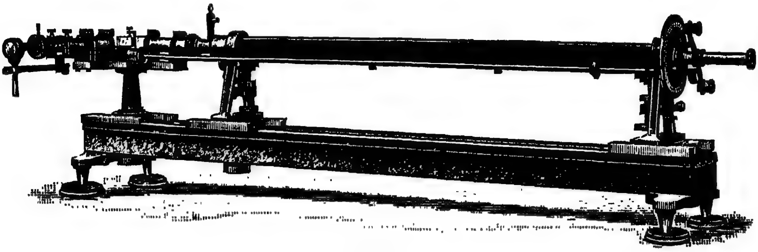


FIG. 24.

ticular wave-length that is employed; thus a set of three interchangeable prisms giving direct vision for the red, green, and violet respectively are required for a complete examination of the rotation of a substance over the whole visible spectrum.

In a spectro-polarimeter of Bellingham and Stanley, the light source is placed directly above the prism (no collimator and slit being used), whilst in the Schmidt and Haensch instrument a direct vision spectroscope is mounted with a second slit, at the focal plane of the condenser in the polariser head.

§ (15) POLARIMETERS FOR SPECIAL PURPOSES. (i.) *Micro-polarimeter.*—When it is necessary to find the rotation of substances that can only be obtained in small quantities, diaphragms are mounted at both F and J (*Fig.* 18) so as to limit the beam. The tube into which the substance is placed has often to be a capillary tube of about 1 mm. diameter. In the instrument made by Schmidt and Haensch the diaphragms at F and J have, therefore, apertures slightly less than this, in order to avoid possible reflections from the walls of the tube. As the field of view is so small (it is not practicable greatly to increase the magnification of the observing telescope without making the field too faint for accurate setting) it is best to use a two-field Lippich polariser, the separating line of the half Lippich prism always bisecting the field whatever size diaphragms are used. This instrument is fitted with a direct vision spectroscope (as described above), with a Nernst lamp attached as the light source. The small aperture diaphragms are removable so that the instrument can also be used for ordinary polarimetry, where the supply of a substance is sufficient for the usual observation tubes.

(ii.) *Diabetometer.*—When the total rotation of the substance is small it is possible to dispense with the circle and yet be able to measure a rotation accurately. In the diabetometer of Yvon, as made by Ph. Pellin (Paris), the analysing nicol is mounted in the tube A (*Fig.* 25). When the screw D is turned by means of the divided drum E, the part B, being engaged to the screw, is rotated about the centre of A. The block C is merely to prevent E from turning too easily, as the screw will have to move this backwards and forwards against the frictional pressure of the spring.

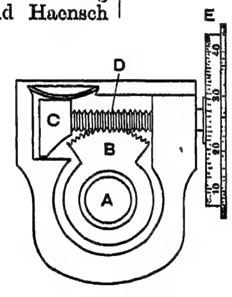


FIG. 25.

The scale on the drum is graduated so as to give directly the number of grammes per litre of sugar or glucose in diabetic urine, allowance being made for the addition of 10 per cent (by volume) of subacetate of lead needed for clarification.

§ (16) LIGHT SOURCES FOR POLARIMETRIC WORK.—As sources of sodium light a very large number of lamps have been designed. For most purposes it is sufficient to use a piece of fused borax on the grid of a Meker-Bunsen burner; in many commercial lamps, the salt is contained in a platinum boat, and especially when a coal gas-oxygen flame is used, the intensity is considerably greater. The Bureau of Standards found[1] that for an intense sodium light source it is best to feed some form of fused Na_2CO_3 into an oxy-hydrogen flame. The borax must not be allowed to melt too quickly, otherwise a reversal of the lines is obtained. For precision work, the Na lines must be regarded as an inferior source, since, although the separation of the lines (6 Å.) is too great to consider the doublet as a homo-

[1] *Circl. Bur. Standards*, 1918 (2nd ed.), No. 44, p. 15.

geneous source, yet the lines are too close for convenient spectroscopic separation.

The green line of mercury (5461 Å.) forms a very suitable light source for spectro-polarimetric work. Although this line has a complex structure when analysed by means of an echelon grating or other high resolving apparatus, the difference in wave-length between its extreme satellites is less than ·4 Å., fifteen times less than the separation of the two sodium lines, and it is found that it can be regarded as homogeneous for rotations as great as 250°. If a quartz mercury-vapour lamp of the type manufactured by the Westinghouse Cooper-Hewitt Co. is used, this source is of great intensity, and can be used indefinitely with little or no attention; although its radiations are, owing to the higher pressure, not so homogeneous as those of the glass tube lamps of the same company. The violet line of mercury, though not so prominent, can be successfully employed; the yellow lines cannot be so readily used, since here we are confronted with the same difficulty as in the case of the sodium doublet, a separation of only 21 Å. The most useful lines are therefore 5461 (green) and 4359 blue (the blue-green 4916 is only about half the intensity of the other lines and does not prove very satisfactory).

Lowry [1] has suggested that, in addition to the sources and lines mentioned above, the following lines are useful for spectro-polarimetry: the lithium red 6708 and the cadmium lines 6438 (red) and 5086 (green). The lithium line is obtained as in the case of the sodium, a small oxy-hydrogen flame being used if great brightness is required, while the cadmium lines are given by a rotating arc. The electrodes are made of 28 per cent cadmium and 72 per cent silver, and are rotated in opposite directions at a speed sufficiently high to prevent flickering. A more convenient cadmium source is the cadmium vapour lamp of Sand,[2] a modification of which has been suggested by Bates,[3] but unfortunately no reliable cadmium vapour lamp has yet been put on the market.

§ (17) Observation Tubes.—The simplest form of polarimeter tube is shown in the left-hand part of *Fig.* 26. It consists of a straight tube of glass of an internal diameter of about 10 mm. and with walls about 2 mm. thick. The ends are ground to the correct length, and should be parallel to within a minute of arc. The ring C, the outside of which is threaded, is cemented on the glass tube, so that the latter projects slightly. The end cap D screws on the ring C, and this carries the end plate or cover glass G. In order that the plate should be held evenly against the end of the glass tube, a soft rubber washer R is placed between the plate and the face of the end cap.

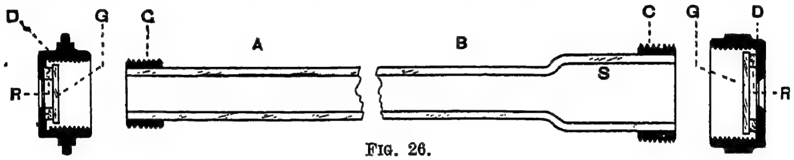


Fig. 26.

Unless the tube is very carefully filled a small bubble of air is left in the tube, which often materially obstructs the path of the beam. This difficulty was overcome by Schmidt and Haensch (Berlin) by enlarging one end of the tube as shown on the right-hand side of *Fig.* 26. The bubble rises to the top of the enlargement S, which is outside the direct path of the beam. The same purpose is effected in the tubes of Bellingham and Stanley (London) by blowing a small bubble or enlargement at the centre of a straight tube.

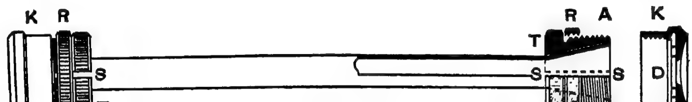


Fig. 27.

The cement with which C is fastened to the tube often becomes loose in time, and the screwing of the end cap tends to draw the ring over the end of the tube, virtually lengthening the column of liquid and so introducing an error in the rotation per unit length. In order to avoid this Neumann [4] has designed the tube shown in *Fig.* 27. Both ends of the tube are ground conically. The ring T is ground to fit the cones of the tube exactly, and is placed in position by splitting the rings along SS. Another ring R is screwed up as shown and serves to keep the split ring tightly clamped to the cone of the tube. The ends A are then ground to the correct length and parallel, the end cap K and the cover glass

[1] Lowry, *Phil. Mag.*, 1909, xviii. 320.
[2] Sand, *Proc. Phys. Soc.*, 1915–16, xxviii. 94.
[3] Bates, *Sci. Paper Bur. Sta.*, 1920, No. 371.
[4] Neumann, *Chem. Zeitung*, 1913, No. 51.

D being the same as in the ordinary tube. The great advantage of this type of tube is that it can be used with ether, alcohol, chloroform, benzol, zylol, etc., which attack the cement in many of the ordinary tubes. The tube can be made of fused silica or porcelain and used for high temperature work in an electric furnace, when the cement of the ordinary tube would melt. It is made by Schmidt & Haensch (Berlin).

In very accurate work, it is necessary that the temperature of the substance should be uniform and that it should be accurately known, since rotatory power is a function of the temperature. The most convenient way of doing this is to surround the tube with another metal tube having small side tubes in it that can be connected to a thermostat. The better made water-jacketed tubes, as they are called, have baffles mounted between the outer and inner tube, so that the flow of the water should be helical and envelop the inner tube. An opening in the middle of the inner tube (see *Fig.* 28) provides for the insertion of a sensitive thermometer, the rotation not being measured until the temperature has become steady.

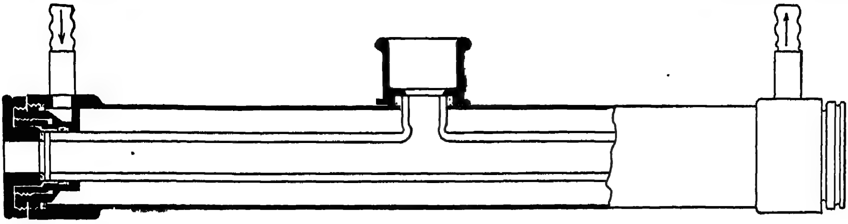


FIG. 28.

When rotations have to be determined for temperatures below the dewpoint of the atmosphere, the end plates must be protected against the condensation of moisture. Wiley [1] employed a second cap made of some non-conducting material such as hard rubber; the air-space in between the two cover glasses contain some desiccating substance.

The jacketed tube is not very satisfactory for use at comparatively high temperatures owing to the fall of temperature along the tube due to conduction, etc. It is then preferable to have the tube in a thermostat mounted between the polariser and analyser.[2]

The measurement of rotation in fermenting liquids (for biochemical work) has to be carried out at 37° C. Abderhalden,[3] in conjunction with Schmidt and Haensch, has evolved an ingenious arrangement whereby six small tubes are mounted on a horizontal wheel in a circular box, any one of which can be brought in line with the optical axis of the polarimeter. The tubes are electrically heated and the temperature can be kept constant at 37° C. by means of a rheostat.

[1] Wiley, *Journ. Amer. Chem. Soc.*, 1896, xviii. 81.
[2] Paul, *Zeitschr. phys. Chem.*, 1916, xci. 745.
[3] Abderhalden, *Hoppe-Seyler's Zeits. für physiolog. Chem.*, 1913, lxxxiv. 300.

§ (18) APPLICATIONS OF POLARIMETRY.—The property of optical rotation is of prime importance in organic chemistry. The work of Pasteur (1848) showed that substances otherwise identical can exist in a right-rotating, left-rotating, and inactive form. Later he introduced the conception of molecular assymetry, which led van't Hoff and Le Bel (both in 1874) to the discovery that optical activity always indicated assymetric distribution of the carbon valencies in the molecule. The further developments of this theory are fully dealt with in text-books on stereochemistry. Wilhelmy in 1850 used the polarimeter to measure the rate of inversion of cane sugar, and employed the results as a basis for the first correct mathematical treatment of the velocity of a chemical reaction.

The polarimeter is extensively used in analytical chemistry for quantitative determinations. In this connection, the term *specific rotation*, introduced by Biot, is employed. By this term is meant the angle of rotation caused by a length of 10 cm. of a solution containing 1 grm. of active substance per c.c. This will vary with the wave-length used and with the temperature, and is designated by the abbreviation $[\alpha]_D^{20°}$, which indicates the specific rotation of a substance at 20° C. using the yellow (D) lines of sodium as a light source.

It then follows that $[\alpha]_D = a/l \times c$, where a is the observed rotation and c the number of grammes of the substance per c.c. of solution. $[\alpha]$ varies not only with the wave-length and temperature, but also with the concentration, the relationship being somewhat complex; fortunately the variation is usually small, and need only be taken into consideration in accurate work such as the determination of sugars. Knowing the specific rotation, the polarimeter can be used for a large number of analytical determinations, such as the estimation of the alkaloids, camphors, and essential oils. Rosin

or castor oil may be detected in mineral oils by their optical activity.

The use of the instrument is not necessarily confined to the determination of optically active substances; a large number of inert substances exert a powerful influence on the rotatory power of active solutions. Thus acetone or boric acid can be determined in solutions of tartaric acid of known strength. The solvent used to dissolve an optically active substance has considerable influence on the rotatory power of the latter, and this property will very likely be capable of many new analytical applications in the near future.

The instrument forms a useful tool for the study of fermentation and other problems in biochemistry.

W. E. W.
F. T.

Polarisation, Plane of: the plane of incidence of a polarised ray upon a reflecting surface when the reflected ray is of maximum intensity. According to the electromagnetic theory of light, the plane of polarisation is at right angles to the direction of the electric force and parallel to the direction of the magnetic force in the wave front. See "Polarised Light and its Applications," § (2).

Polarisation of Light, Plane, Circular, and Elliptical. See "Polarised Light and its Applications."

Polarisation Photometers. See "Photometry and Illumination," § (30).

Polarisation Spectrophotometers: instruments in which the brightness match is obtained by the rotation of a polarisation device. See "Spectrophotometry," § (12).

Polariscope, Tourmaline. See "Polarised Light and its Applications," § (9).

POLARISED LIGHT AND ITS APPLICATIONS

§ (1) Introduction.—Fresnel's original conception of the disturbance producing the sensation of light was, briefly, a train of waves propagated by vibrations of ether particles taking place in all directions perpendicular to the direction of propagation of the disturbance. Later, he modified this conception to obviate the mechanical objection to two adjacent particles moving in totally different directions, and he postulated a vibration at any given moment of all particles along parallel lines, but with so rapid a change in the common direction of motion that a period of vibration in any one direction cannot be isolated in the analysis of a beam of ordinary light. Under special conditions, however, the vibrations may be restricted in such a way that any one particle continues to describe the same path, whether it be linear, circular, or elliptical, and all the particles affected by the disturbance describe exactly similar paths. It is this restriction of vibration that is known as the polarisation of light.

§ (2) Discovery of Fundamental Phenomena.—Two and a half centuries ago Erasmus Bartolinus discovered the power of a crystal of Iceland spar to resolve a beam of ordinary light into two separate rays; twenty years later Huygens repeated his work, and found that two rays were produced unless the light was travelling parallel to the crystallographic axis of the crystal, and, moreover, that each ray could again be subdivided by its passage through a second crystal unless the latter was orientated in one of four definite directions relative to the first crystal. In that case the two rays were transmitted without further division. A few years later Newton confirmed these observations, but still no explanation was forthcoming, although both these pioneers recognised the fundamental property of polarised light, namely, its asymmetry about its direction of propagation. "Have not," wrote Newton, "the rays of light several sides endued with several original properties?"

Early in the nineteenth century Malus made the chance discovery that light reflected from glass possessed in part the characteristics of rays emerging from a calcite crystal, and that these characteristics became most marked for a definite angle of reflection. Experimenting later with light reflected from other surfaces, he found that this held good for all reflecting surfaces other than metals, but that the angle producing the maximum result varied with each individual substance. Endeavouring to find an explanation for these facts, he rejected the wave theory of light as offering no possible solution: at that time Young's longitudinal vibration theory had not yet been superseded by Fresnel's more far-reaching transverse vibration theory, and obviously the former could give an even less satisfactory explanation than could the older corpuscular theory of Newton. Malus concluded that the phenomena must be due to some induced property of the corpuscles transmitting light rays, comparable with a magnetic or electric charge, and producing in the light a definite bias or "polarity," and this gave birth to the title of "polarisation," which has survived the destruction of the theory to which it owes its origin. The group of workers responsible for leading the explanation along new theoretical lines included Young, Fresnel, Arago and Brewster, who suggested and elaborated a theory of polarisation based on a wave theory of light. Nevertheless,

although his efforts at explanation did not stand the test of time, Malus accumulated a mass of results from minute and accurate observation on the behaviour of polarised light which furnished conclusive evidence for the law which bears his name.

The asymmetry of a ray polarised by reflection becomes very evident when the ray suffers reflection from a second surface, and Malus found that the intensity of the final ray was dependent upon the plane of incidence at the second surface. He attributed to each ray a plane passing through its path which bore a definite relation to its asymmetric properties; this he called the *Plane of Polarisation* of the ray, and defined it as the plane of incidence of a polarised ray upon a surface when the reflected ray was of maximum intensity. The angle of reflection from a surface for which the polarisation of the light was most complete he termed the *Angle of Polarisation.* His *Law of Polarisation* could then be stated as follows:

When a ray of light polarised by reflection at one surface falls upon a second surface at the angle of polarisation, the intensity of the twice-reflected ray varies as the square of the cosine of the angle between the two planes of reflection.

The foundations of the theory of polarisation which holds to the present time were laid by Young and Fresnel in their conception of light transmission by transverse wave motion. Fresnel saw the immediate application of the conception to the case of polarised light, and suggested the restriction of the vibrations to a definite unchanging path at right angles to the direction of propagation. Accepting his hypothesis, we shall for the moment consider only light whose vibrations are linear and take place in one definite direction. To this has been given the name of *Plane Polarised Light.*

A crucial test of Fresnel's theory lay in the investigation of the behaviour of two plane polarised rays under conditions which might be expected to produce interference. Fresnel and Arago carried out this work, and their results added evidence in favour of the wave motion hypothesis, and led to the establishment of the following five laws governing the interference of polarised rays, known as the *Fresnel-Arago Laws*:

1. Two rays polarised in the same plane interfere with each other under the same conditions as for ordinary light.
2. Two rays polarised at right angles to one another do not interfere under these conditions.
3. Two rays polarised at right angles, if obtained from unpolarised light, may subsequently be brought into the same plane of polarisation without acquiring the power of interference.
4. Two such rays derived from plane polarised light will, under the same conditions, show interference.
5. In the latter case a phase difference of π, equivalent to half a wave-length, must be added to the estimation of the path difference.

At about the same time Brewster deduced from experimental data another law governing the size of the angle of polarisation. *Brewster's Law* may be expressed simply in the statement that the tangent of the angle of polarisation is equal to the index of refraction of the reflecting substance. From this it follows that at the polarising angle the sum of the angles of incidence and refraction is 90°, or that the reflected ray is perpendicular to the refracted ray, and that when light travelling in a medium is polarised by reflection from the bounding air-surface of the medium the refractive index is the cotangent of the polarising angle. Evidently, therefore, the polarising angle will vary with the wave-length of the light used. With most substances the dispersion is too low to show the effects of this law, but with substances of very high dispersive power, such as nitroso-dimethyl-aniline, the effect is visible in the distinct coloration of the image of a white source of light, and in the variation of the colour with the angle of incidence.

By the examination of light polarised by refraction through a crystal or by reflection at a glass surface various definite facts were deduced. The two rays emerging from a crystal of calcite were found to be polarised with their planes of polarisation parallel and perpendicular respectively to the plane containing the incident ray and the crystallographic axis. The ray reflected from a polished non-metallic surface was polarised in the plane of incidence and the refracted portion was partially polarised in a perpendicular plane. After a second refraction at the polarising angle the percentage of polarised light in the refracted ray was increased, while again the reflected ray was wholly polarised in a perpendicular plane, and this process could be repeated at any number of surfaces until the final refracted beam contained no appreciable amount of unpolarised light. These facts are capable of simple explanation on the wave theory of light propagation.

§ (3) Explanation on Wave Theory.—A transverse vibration in any direction may be resolved into two component vibrations at right angles in the same plane; moreover, it can be shown that an elliptical or circular vibration can be resolved into two simultaneous linear vibrations at right angles to each other differing in phase. It may be supposed that a crystal such as calcite has an inherent power of resolving light vibrations in this way; a separation of the two rays would be effected if we suppose that the one set of vibrations possesses a different rate of propagation through the crystal from that of the perpendicular set. This subject will be investigated more fully a little later in the general discussion of the behaviour of crystals in transmitting light. Turning now to the question

of polarisation by reflection, the wave theory supplies an equally simple explanation. When a beam of light meets a flat surface of a transparent medium, part of the light is reflected, but the greater part is transmitted. If in a ray of light incident on a surface at the polarising angle the vibrations are resolved into directions parallel and perpendicular to the plane of incidence, the polarised reflected ray can be regarded as derived from the reflection of part of the component whose vibrations are perpendicular to this plane. No part of the component in the plane appears in the reflected ray. The refracted light then consists of the remainder of the perpendicular component together with all the light vibrating in the plane of incidence, and the preponderance of vibrations in the latter direction causes the refracted beam to appear partially polarised in the corresponding sense. On a second reflection the degree of polarisation of the refracted ray is increased by a repetition of the same process, while the reflected light again consists of vibrations perpendicular to the plane of incidence, and reinforces the beam reflected at the first surface. Each additional surface augments this double process of reflection and refraction until the final refracted ray is deprived of all vibrations perpendicular to the plane of incidence, and consists wholly of vibrations taking place in this plane.

§ (4) Relation between Plane of Polarisation and Direction of Vibration.—So far no definite relationship has been established as existing between the plane of polarisation, as defined by Malus, and the plane in which the vibrations take place. Various investigators have assumed the alternate possibilities of the parallelism or perpendicularity of these planes. Fresnel, to whom we owe so much of our knowledge in this branch of Optics, assumed as a basis for his mathematical investigations that the vibrations were perpendicular to the plane of polarisation: M'Cullagh, on the other hand, deduced as a consequence of his theory that the vibrations were in Malus's plane of polarisation. At the present time the electromagnetic theory indicates the existence of vibratory motion in both planes; the electric force is perpendicular to the plane of polarisation, the magnetic force is in that plane, and both these are concerned in the transmission of the light.

Stokes, working on the elastic solid theory, produced his Dynamical Theory of Diffraction,[1] wherein he arrived at a result which could be employed as a criterion of the actual vibration direction in plane polarised light. He showed theoretically that, if plane polarised light is diffracted, each diffracted ray is also plane polarised, and further, that if the plane of polarisation of the incident ray is inclined successively at regularly increasing angles to the plane of diffraction, the planes of polarisation of the corresponding diffracted rays are crowded together towards the plane of diffraction or towards a perpendicular plane according as the vibrations are parallel or perpendicular to the plane of polarisation. Experiment with finely ruled gratings showed that the latter effect is produced, which indicates the correctness of Fresnel's hypothesis.

Another suggestive experiment was carried out by Tyndall. A ray of polarised light was passed longitudinally through a tube containing very fine particles in suspension. If the vibrations of the beam were confined to one plane, say the vertical plane, the beam would possess no energy capable of producing horizontal vibrations. It has been shown that the intensity of plane polarised light scattered by fine particles is zero in the direction parallel to the vibrations of the light; consequently, there would in this case be no scattered light visible in the vertical direction; the vertical vibrations would, however, render scattered light visible in the horizontal direction perpendicular to the axis of the tube. Using light polarised in the horizontal plane, Tyndall found that scattered light was visible only in a horizontal direction, which may be taken as an indication that the vibrations in a plane polarised ray are perpendicular to its plane of polarisation.

§ (5) Double Refraction.—It will be convenient at this point to consider the general behaviour of crystals in transmitting light, as upon this we are dependent to a great extent for the production, analysis, and technical applications of polarised light.

A substance in the crystalline state differs from an ordinary isotropic medium in that many of its properties may vary in different directions. All crystals can be referred to definite systems of crystallographic axes,[2] and are divided into classes according to the relative length and inclination of these axes. In its natural form the shape of a crystal is governed by the character of the axes, which determine the size of the angles contained by all possible crystal faces, and just as the morphology of a crystal varies along the different axial directions, so there is a corresponding variation in other physical properties, mechanical, thermal, electrical, magnetic, and optical. In short, a crystal is possessed of a definite molecular structure on which depend the physical properties along any definite direction. In the most regular system, the cubic, the three axes are all equal and perpendicular, and the optical properties are constant in all directions, but in other systems, as the relation between the crystallographic axes becomes more complex and less regular, corresponding variations appear in all physical properties.

It has already been noted that the property of double refraction in Iceland spar was discovered as early as 1670, and later Huygens and others established definite facts connected

[1] *Mathematical and Physical Papers*, 1901, ii.

[2] See "Crystallography," § (4), etc.

with this phenomenon. A ray of light in passing through a calcite crystal is in general divided into two perpendicularly polarised rays, only one of which obeys the laws of ordinary refraction. They are termed respectively the *ordinary* and *extraordinary rays*, and they travel along slightly different paths, the separation between them varying with their inclination to the crystallographic axes, and becoming zero when the light travels along a definite direction in the crystal known as the *Optic Axis*. All doubly refracting crystals may be classified as *uniaxial* or *biaxial* according to the presence in the crystal of one or of two directions along which there is no separation of a refracted ray. Since the case of uniaxial crystals is very much simpler, we shall deal first with that alone; moreover, this class is of the greater importance in optical work, as it includes the two crystalline media most commonly used, namely, quartz and calcite.

§ (6) Uniaxial Crystals.—The separation of the rays produced by refraction indicates the existence of two values of the refractive index of the crystal for rays polarised in two perpendicular planes. If the direction of the incident ray is varied it is found that the refractive index corresponding to the ordinary ray is constant for all directions through the crystal, while that for the extraordinary ray varies, reaching extreme values parallel to the optic axis, when the index is that of the ordinary ray, and perpendicular to the axis. It can be shown, too, that the ordinary ray is always polarised in the principal plane, that is to say, in the plane containing the refracted ray and the optic axis, so that its vibrations are always perpendicular to the optic axis, while the extraordinary ray is polarised in the plane at right angles to the principal plane.

(i.) *Huygens' Construction for Path of Rays.*—Although with his conception of the nature of light transmission Huygens could not explain the polarisation of doubly refracted rays, he was able to represent the formation and paths of the two rays by a graphic method,[1] attacking the problem by a consideration of the wave-fronts of light disturbances travelling in a crystalline medium. Supposing a disturbance to emanate in all directions from an isolated point in the crystal, two separate wave-surfaces will be formed, that of the ordinary rays travelling at the same speed in all directions, and spreading out in the form of a sphere with ever-increasing radius. The rate of propagation of the extraordinary ray will vary along different directions, and consequently the wave-surface will no longer be spherical. Huygens assumed that it would take the form of a spheroid generated by the revolution of an ellipse, the axis of revolution coinciding with the diameter of the spherical wave-surface that is parallel to the optic axis. This assumption satisfied the necessary condition that the refractive index for the extraordinary ray should have an extreme value in the direction of the optic axis, and a second extreme value in all directions perpendicular to the axis.

Huygens' graphical method of determining the position of a refracted plane wave-surface by considering each point on the refracting surface as an origin of secondary disturbance is well known. Suppose AB (*Fig.* 1) represents

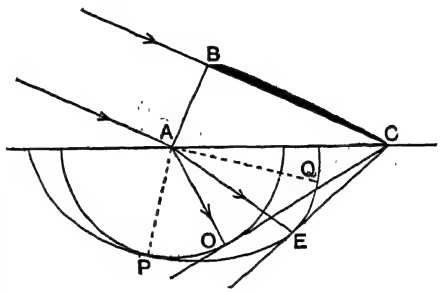


Fig. 1.

a section of the plane wave-front impinging on the plane surface AC of a doubly refracting medium, both planes being perpendicular to that of the paper. The point A may be taken as an origin for the two disturbances propagated in the crystal, one of which will spread in a spherical form, and the other in a spheroidal form. Let the plane ABC contain the optic axis, which lies in the direction AP; then these wave-surfaces may be represented by the circle through P and O with its centre at A, and the ellipse through P and E with AP and AQ as its axes, where AP/BC and AQ/BC represent the extreme values of the refractive index, the former for both rays travelling along the optic axis, and the latter for the extraordinary ray travelling perpendicular to it. The refracted wave-fronts of the whole beam will consequently be represented by the tangent planes[2] respectively to the ordinary and extraordinary wave-fronts. These will be perpendicular to the paper and cut it in lines CO and CE respectively. It is clear that in the case of the latter ray the sines of the angles of incidence and refraction bear no constant relation to each other, and, moreover, the wave-front CE is not perpendicular to the direction of the ray AE unless the latter is travelling parallel or perpendicular to the optic axis.

If the optic axis is not co-planar with AB and AC, the plane through C and the intersec-

[1] See "Light, Double Refraction of."

[2] See "Light, Propagation of."

tion of the incident wave and the refractory surface tangential to the spheroid described round A will not touch the spheroid in the plane of incidence ABC; in this case, therefore, the extraordinary ray does not obey either of the laws governing ordinary single refraction. Various special cases can be examined by applying the necessary modification to the graphical representation of the simple case considered here.

Much work has been done by later investigators, proving that Huygens was justified in his assumption as to the shape of the wave-surface of the extraordinary ray. There are two possible cases to be considered: the refractive index for the extraordinary ray perpendicular to the optic axis may be either less or greater than that for the ordinary ray. In the former case the wave-surfaces within the crystal will take the form of a spheroid enclosing a sphere, having a common diameter parallel to the optic axis; calcite belongs to this class of crystals, which are said to possess *negative birefringence*, or, more shortly, to be negative crystals. In the other case the spherical wave-surface encloses the spheroid, and the crystals are said to be *positive*. Quartz may be instanced as an example of a positive crystal, although it will appear later that owing to its rotatory power the two wave-surfaces in a quartz crystal do not touch at any point, that is to say, the difference in refractive index for the two rays reaches a minimum along the optic axis, but never disappears entirely.

§ (7) Biaxial Crystals. (i.) *Wave Surface.*—Huygens did not extend his theory to explain the phenomena exhibited by biaxial crystals, and his construction is not applicable to this case. It is to Fresnel that we owe the earliest complete theory of double refraction, dealing with both biaxial and uniaxal crystals. Some account of this will be found in the article "Light, Double Refraction of." Meanwhile, it is sufficient to say that the wave-surface is of a somewhat complicated form, having two sheets; as shown in *Fig.* 2 the section of the surface by each of the principal planes is a circle and an ellipse.

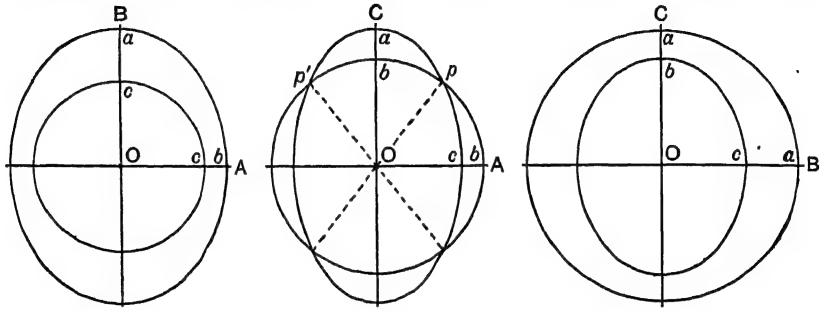


Fig. 2.

Fletcher's Indicatrix.—Another widely used representation of the optical properties of a crystal is *Fletcher's Indicatrix*. It is identical with the older "Index-ellipsoid" of M'Cullagh, the ellipsoid being constructed on semi-axes proportional to the three refractive indices referred to before (article "Light, Double Refraction of"). Space does not permit of a full explanation of the properties of the Indicatrix, which may be found in the larger treatises on Crystallography, or in the original paper by its inventor;[1] it must suffice to state that in any section of the ellipsoid the axes are proportional to the refractive indices of the two polarised rays whose wave-fronts move along the normal to that section, and indicate the vibration directions in the rays, while their difference in length gives a measure of the double refraction along the normal to the section.

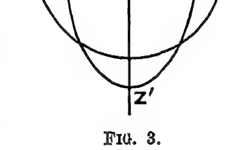


Fig. 3.

(ii.) *Primary and Secondary Optic Axes.*—The position of a plane wave for which there is only single refraction through the crystal will, as we have seen (article "Light, Double Refraction of"), be the plane tangential to the two wave-surfaces. Taking a section in that plane through two axes of the ellipsoid in which two sheets of the wave-surface intersect in N (*Fig.* 3), MM' represents the common tangent which touches the circular section in M and the elliptical section in M'. OM, the normal to the tangent plane through M, represents the direction in which a single wave-front is propagated, and, therefore, is the optic axis. ON, joining the centre to the point of intersection of the circle and ellipse, and representing the direction in which a single ray is propagated, is sometimes described as the *Secondary Optic Axis*; or it is described more completely by the name *Axis of Single Ray Velocity*.

[1] *Min. Mag.*, 1891, ix. 278.

(iii.) *Conical Refraction.*—Now if the complete wave-surfaces are considered and not merely the sections in the ZX plane, it can be proved mathematically that the common tangent plane through MM′ touches the surfaces in a circle and not merely in two points, as Fresnel originally assumed.

If, then, a plane wave is incident on the crystal parallel to the tangent plane MM′, any line joining points on this circle of intersection to the centre O will be possible directions for rays entering the crystal. The direction will be determined by the plane of polarisation of the incident light; if the latter is unpolarised, that is to say, if it contains vibrations in all possible directions, the refracted rays will take the form of a hollow cone of rays, each element of the cone being polarised in a separate plane. This phenomenon, known as *internal conical refraction,* was predicted by Sir William Hamilton,[1] and verified practically by Dr. Lloyd.[2] An analogous effect, known as *external conical refraction,* is produced by the refraction out of a crystal of a ray travelling along ON, or the axis of single ray velocity. The direction of the wave normal in the crystal will be normal to the plane touching the wave-surface at the point where it is intersected by the ray inside the crystal. Since the ray ON meets the wave-surfaces at the apex of a cone, there are an infinite number of tangential planes to the wave-surfaces at this point, and the normals from N to these planes will form a hollow cone. The ray ON when refracted out of the crystal will therefore form a hollow cone of light. The practical demonstration of both these phenomena, which are consequent on the theory of birefringence in crystals, helps to establish the correctness of that theory.

(iv.) *Relation between Morphological and Optical Properties.*—It is essential from the point of view of practical work with actual crystals to have a clear idea of the relation between their morphology and their optical properties.[3] It has already been stated that crystals exhibiting the highest possible symmetry, and therefore belonging to the cubic system, possess identical optical properties in all directions, and from an optical point of view behave in all respects as an isomorphous medium. Crystals of the tetragonal, hexagonal, and trigonal systems are uniaxial, their optic axis coinciding with the main crystallographic axis. Crystals of other systems, the rhombic, monoclinic, and triclinic systems, are biaxial. In a rhombic crystal the axes of Fresnel's or of Fletcher's ellipsoid coincide with the crystallographic axes in direction, although not necessarily in length. A monoclinic crystal possesses one crystallographic axis perpendicular to the plane containing the other two, which may be inclined to each other at any angle, and this axis, about which the crystal is symmetrical, coincides with one axis of the ellipsoid. The other two axes of the ellipsoid lie in the perpendicular plane of symmetry, but neither of necessity coincides with a crystallographic axis, and their position may show changes with a change of temperature or wave-length. In a triclinic crystal, which possesses no symmetry about any crystallographic axis, there is no definite relationship between the positions of the morphological and optical axes.

[1] *Trans. Roy. Irish Acad.*, 1833, xvii.
[2] *Phil. Mag.*, 1833, ii. 112 and 207.
[3] See "Crystallography," §§ (7) and (21).

Dispersion of Axes and Bisectrices.—In dealing with the position of the optic axes and the bisectrices of the optic axial angle it was assumed that the wave-length of light remained constant. If white light is used in the examination of a crystal there may be different directions in the crystal for different wave-lengths along which there is no double refraction, and this will cause a *dispersion of the optic axes. Dispersion of the bisectrices* also may take place, the effects varying in different classes of crystals. In all cases in which an optic axis or a bisectrix coincides with a crystallographic axis there can be no dispersion; hence in uniaxial crystals the position of the axis is independent of wave-length. In rhombic crystals the bisectrices coincide with crystallographic axes, so that dispersion is limited to the optic axes. In some crystals it is almost too small for measurement, but in extreme cases the optic axial angle for red light is large, decreasing with decrease of wave-length until the crystal becomes uniaxial for light of a definite wave-length; for still shorter wave-lengths the optic axial angle opens out again but in the perpendicular direction. This phenomenon is known as *crossed axial dispersion.*

In monoclinic crystals dispersion of the bisectrices may also occur, since of the ellipsoidal axes only one coincides with a crystallographic axis. The effect produced varies according as this crystallographic axis coincides with the acute bisectrix or the obtuse bisectrix, or is perpendicular to the optic axial plane. Dispersion of the optic axis may also appear simultaneously and in certain cases may be unequal for the two axes. For fuller information on this subject the reader is referred to the standard text-books on Crystallography and Petrographic Methods by Tutton, Miers, Johannsen, Rosenbusch, Weinschenk and Clark, and other writers.

§ (8) Methods of Production of Polarised Light.—The various methods of producing a ray of plane polarised light from an ordinary unpolarised beam all depend fundamentally upon the resolution of the light into two components with vibrations in perpendicular directions, combined with the isolation of one component.

(i.) *By Reflection.*—We have seen that this can be effected very simply by repeated refractions of a ray incident at the polarising angle. A pile of thin glass plates may be

used for the purpose; while an increase in their number tends to increase the degree of polarisation of the transmitted light, it decreases the intensity simultaneously, and to minimise this disadvantage the plates used should be as thin as possible consistent with a fair degree of rigidity. Seven or eight plates of the thickness used for cover-glasses are sufficient to produce satisfactorily complete polarisation of both the reflected and the refracted light, while not reducing too greatly the intensity of the latter.

It is frequently desired to study the behaviour of various substances in polarised light, and generally, as well as a means of polarising the incident light, we require a means of analysing the final transmitted or reflected rays. The same instrument is usually capable of fulfilling both purposes, for if it transmits light vibrating in one direction only, thereby acting as a polariser, it can also be used as an indicator of the vibration direction of a polarised beam, and fulfils the function of an analyser. For some purposes the vibration direction of the analyser is arranged parallel to that of the polariser, but more usually their vibration directions are perpendicular; in the latter case the polariser and analyser are said to be "crossed."

(ii.) *By Refraction.*—For some purposes for which a small readily handled contrivance is preferable to a more perfect instrument requiring careful manipulation, a plate cut from a large tourmaline crystal parallel to the long hexagonal axis is found of use. Tourmaline is strongly birefringent, and also possesses a different coefficient of absorption for the two rays, with the result that the ordinary ray is completely absorbed, while the extraordinary ray is transmitted. Selective absorption lends the transmitted light a deep green colour, however, which greatly reduces its intensity and renders this method of production unsuitable for many purposes.

§ (9) Tourmaline Polariscope.—A simple polariscope can be constructed of two tourmaline plates mounted one above the other, and held in a pair of wire tongs in such a way that either plate can be rotated relatively to the other. The contrivance is well adapted for a rough examination of the behaviour in polarised light of a crystal specimen placed between the two plates. The lower plate acts as a polariser of light passing upwards through it, while the upper plate is used to analyse the light emerging from the specimen. If the plates are so orientated that the vibration direction of the extraordinary ray in the upper is at right angles to that for the lower plate, no light can pass through both plates, unless the presence of a crystal slice between them causes a rotation of the vibration direction of light transmitted by the lower plate. Any such effect can be examined by a rotation of the upper tourmaline.

§ (10) Nicol Prism.—The heavy loss of intensity due to the strong selective absorption of tourmaline is obviated if calcite is used, and this is employed in the construction of the *Nicol Prism,* which is the most generally used instrument for the production and examination of polarised light. With a rhomb of calcite special means have to be adopted to eliminate one of the two polarised rays. Early in the nineteenth century William Nicol [1] devised a means of attaining this result by slicing a rhomb of calcite diagonally and symmetrically through its blunt corners, polishing the cut surfaces and cementing them together with a thin layer of Canada balsam. The success of the device depends upon the fact that the refractive index for Canada balsam lies between that for the ordinary ray in calcite and the maximum refractive index for the extraordinary ray. Thus the dimensions of the prism must be so designed that the ordinary ray falls on the balsam film at an angle greater than the critical angle, and is totally reflected, while the extraordinary ray is transmitted. The ratio of the long edge of the crystal to the short end-face should be between 1 : 3·0 and 1 : 3·7; the sides of the prism are blackened to absorb the reflected light and the end-faces are polished. This represents the Nicol prism in its simplest form; modifications will be treated of later.

On looking through a Nicol prism and varying the angle of obliquity the observer sees that the polarised field is limited on one side by a blue haze which gradually obscures the light, and on the other by a narrow band of interference fringes beyond which the intensity is increased but all images are doubled.[2] A consideration of the actual passage of rays through the Nicol will show that these limits mark on the one hand the beginning of the region of total reflection of both rays, and on the other the end of the region of total reflection of the ordinary ray, the fringes being produced by the interference of the two rays within the balsam film; in the region beyond the fringes the increase in their path separation causes the formation of double images.

Fig. 4 represents a section of the prism perpendicular to the cut section. In the natural crystal the three faces bounding the obtuse solid angle A (*Fig.* 5) make equal angles with one another and with the crystallographic axis, which coincides with the optic axis AO. This latter lies in the plane ABCD, and makes an angle of about 45° with the end-face AD, the whole angle DAB being about 109°. The

[1] *Edinburgh New Phil. Journ.*, 1833, vi. 182, and 1839, xxvii. 332.
[2] Thompson, *Phys. Soc. Proc.*, 1877, ii. 185.

plane of section AC is inclined at very nearly 90° to the end-face AD. If the transparent film is of Canada balsam, the critical angle

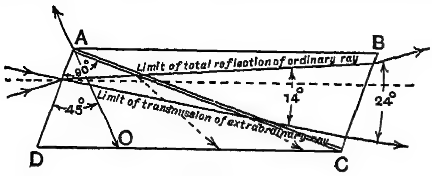


FIG. 4.

for the ordinary ray passing from calcite to balsam is 67° 53′; this limits the polarised field to an angle of 3° from the longitudinal axis in the direction of AB. Now although the maximum refractive index for the extra-

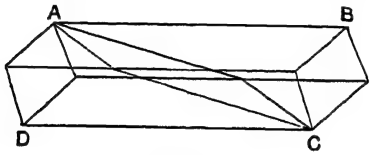


FIG. 5.

ordinary ray is greater than that of the balsam, as the angle of incidence with AC increases, the ray is travelling more nearly parallel to the optic axis, and its index decreases to a smaller value than that of the balsam, so that at a certain angle of incidence the extraordinary ray strikes the balsam at the critical angle. The value of this angle will vary with the wave-length, for the refraction at the end-face AD will cause the red rays to fall on the balsam film at a more oblique angle than the violet rays; hence the red rays are the first to suffer total reflection, and a blue fringe shows the limit to the useful field on the side nearest DC. The fringe appears at an angle of obliquity of about 11° inside the prism; the useful field within the prism is therefore limited to 14°, which corresponds to an external angular field of about 24° in air. The middle line of the field is inclined to the longitudinal axis at about 10°, and this gave rise to the custom often used of grinding the end-faces to make an angle of 68° with the sides before sectioning the prism On the other hand, Ahrens has adopted the practice in late years of trimming the end-faces by about 3° in the other direction, and then cutting the section plane perpendicular to them; this has the effect of increasing the field of view at the cost of making it less symmetrical with regard to the long axis of the prism. Square-ended prisms are also cut, and this not only decreases the percentage of light lost by reflection at the end-surfaces, but obviates any displacement of the image as the prism is rotated about its axis.

Since the extraordinary ray is polarised in a plane perpendicular to the principal plane, that is, perpendicular to the plane ADCB, its vibrations will take place in this plane. The trace of the vibration direction of transmitted light on the end-face of the prism will therefore be along AD, the shorter diagonal of the rhombic face.

§ (11) OTHER POLARISING PRISMS.—Various devices have been adopted to give improvements in the directions of widening the field of view, lessening the percentage of light lost by reflection, increasing the working aperture relative to the length of the crystal, and of giving uniform polarisation over the whole field. To these ends modifications have been made in the inclinations of the section plane and end-faces to the long sides, and in the substance used for the transparent film. A full account of the detailed construction of the Nicol prism, together with a description of other cognate forms of polarising prisms, is given by Silvanus P. Thompson.[1] The chief modifications are due to:

Hartnack,[2] who widened the field of view by cutting the section perpendicular to the optic axis, at a great cost of material.

Foucault,[3] who dispensed with the difficulties attached to cementing the interface by using an air-film. This gave a very narrow field of view and introduced the disadvantage of multiple reflections caused by the air-film.

Glan,[4] who designed a square-ended Foucault prism with the plane of section containing the optic axis.

Jamin,[5] who reversed the construction of the Nicol, and used the ordinary ray transmitted by a thin plate of spar placed between rectangular prisms of carbon bisulphide contained in a metal tube with glass ends. This was further modified by Zenker,[6] who used prisms of dense flint, and by Feussner.[7]

Leiss,[8] who proposed the substitution, for economical reasons, of a prism of glass of similar index for the second half of a Nicol prism.

Thompson (*loc. cit.*) also describes many different forms of prism devised by himself, with references to several other investigators, including Dove,[9] Glazebrook,[10] and Bertrand,[11] who have sought to find the ideal mean

[1] Thompson, *Optical Convention Proc.*, 1905, p. 216.
[2] Hartnack and Prazmowski, *Ann. Chim. et Phys.*, 1866 (4), vii. 181.
[3] *Comptes Rendus*, 1857, xlv. 238.
[4] *Carl's Repertorium*, 1880, xvi. 570; *Journ. de Phys.* x. 175.
[5] *Comptes Rendus*, 1869, xlviii. 221.
[6] *Zeits. Instrumentenk.*, 1885, iv. 135.
[7] *Zeits. Instrumentenk.*, 1884, iv. 49.
[8] *Ber. Akad.*, 1897, p. 901.
[9] *Pogg. Ann.*, 1864, cxxii. 18 and 456.
[10] *Phil. Mag.* (5), xv. 352; *Phys. Soc. Proc.* v. 204.
[11] *Soc. Minéralogique de France Bull.*, 1884, p. 389 *Comptes Rendus*, xcix. 538.

between a prism of greatest utility and one involving small waste of material, an almost equally important consideration owing to the small supply of large flawless crystals of Iceland spar.

Ahrens[1] devised a triple prism (*Fig.* 6), consisting of two calcite prisms cut with their refracting edges parallel to the optic axis, cemented on either side of a calcite prism cut with its edge perpendicular to the axis. A

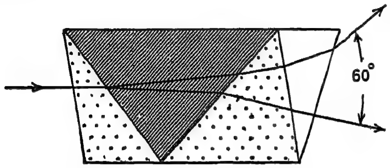


FIG. 6.

rectangular prism of flint glass cemented to the third prism serves to deflect the unused ray 60° from the normal. This prism has a comparatively large angular field. A later form consisted of a triple prism cut from a rectangular block of spar (*Fig.* 7), the three wedges being cemented together with balsam. The ordinary ray is deflected partly to the right and partly to the left, and the extraordinary ray passes through normally, perpendicular to the optic axis. The prism possesses a large aperture relative to its length.

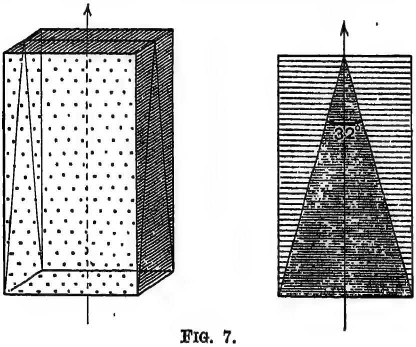


FIG. 7.

Other polarising prisms have been designed by several authors.

Abbe's[2] design consisted of two 30° crown glass prisms cemented with a substance of the same refractive index to the sides of a 60° prism of calcite cut with its refracting edge parallel to the optic axis. The refractive index for the glass must be that for the extraordinary ray in calcite, so that this ray passes undeflected through the block, while the ordinary ray is deviated. The prism has a field of view of about 23°.

Stolze[3] designed a polarising prism constructed entirely of glass. A ray entering normally is twice internally reflected, first at a silvered surface, the polarisation taking place at the second surface by total reflection at the polarising angle, after which the ray emerges normally from the end-face. The lateral displacement of the emergent ray and its incomplete polarisation, due to almost unavoidable strains present in the glass, detract from the utility of this prism.

Schulz's[4] design obviates both these disadvantages, but the prism is not so simply constructed (*Fig.* 8). The surfaces AE and BD are silvered, and a ray entering AB normally leaves EF normally and is reflected at the polarising angle from CE, the emergent beam being in line with the incident beam;

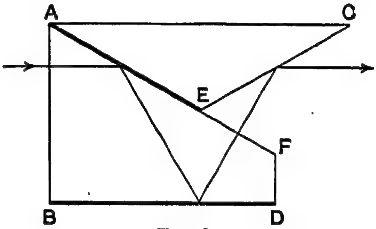


FIG. 8.

the polarisation is unaffected by strain in the glass since it takes place outside the prism. Ten per cent of the incident light is transmitted as compared with 25 per cent to 40 per cent by the Nicol prism, but since no calcite is used there is not the same objection to using a larger prism if a greater amount of light is required.[5]

§ (12) DOUBLE-IMAGE PRISMS.—Various prisms have also been designed to produce two widely separate images of a single object. *Wollaston's Prism* consists of two rectangular prisms of calcite ABC and DAC (*Fig.* 9) cemented together to form a rectangular block. The prism ABC is so cut that the optic axis of the crystal is parallel to the surface AB at which the light enters, while in the other prism the axis is still parallel to the surfaces AB and CD but turned through 90°. A ray of light entering AB normally

[1] *Journ. Roy. Micros. Soc.*, 1884 (2), iv. 533, and 1886, vi. 476; *Phil. Mag.*, 1884 (5), xix. 69, and 1886, xxi. 476.

[2] See Grosse, *Die gebräuchlichen Polarisationsprismen*, 1887.

[3] *Atelier der Photographen*, 1895, p. 140.

[4] *Zeits. Instrumentenk.*, 1911, xxxi. 180.

[5] A device used before the invention of the Nicol prism was the truncation of the blunt corners of a calcite rhomb, each cut limiting the field of view laterally so as to prevent the clear passage of the ordinary ray through the prism. A somewhat similar device has recently been reintroduced for use in polarimetry and other purposes where a wide field of view is not essential; the prisms give good polarisation and are free from the disadvantages attached to a cemented prism.

will travel perpendicularly to the axis and will be divided into two coincident polarised rays vibrating parallel and perpendicular to

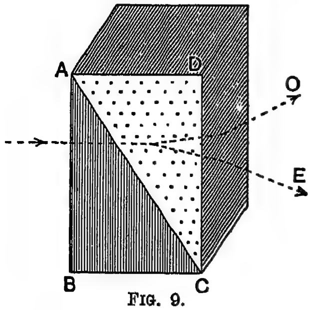


FIG. 9.

the plane of the paper. On crossing the plane of section the ordinary ray of the first prism becomes the extraordinary ray of the second, and *vice versa*; the refractive index therefore increases for one ray and decreases by the same amount for the other ray, so that both are deflected in opposite directions from the normal to AB. The divergence of the rays is further increased on their refraction at the surface CD, and they form two widely separated images polarised in perpendicular planes. On rotating the prism about the normal to AB the two images rotate simultaneously, each retaining the same position relative to the prism. If the prism is tilted so that the rays are incident obliquely the images are seen to separate or approach according to the direction of tilt, always supposing the tilt to be in the plane containing the two images. A modification is sometimes made by using one calcite prism and one glass prism. In this case the change of refractive index for each ray at the interface is less, though still in opposite directions, so that such a prism, although cheaper in construction, does not give as great total deviation.

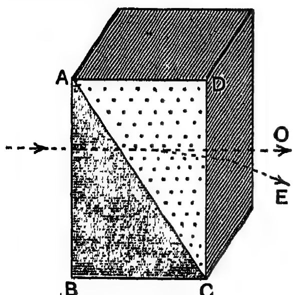


FIG. 10.

Rochon's Prism (*Fig.* 10) is similar in general design, except that the optic axis in the first prism is perpendicular to the entrance face. No separation of normal light occurs in the first prism, therefore, and on passing into the second half the ordinary ray travels on unchanged in direction while the extraordinary ray is deviated further from the normal to the plane of section. On emergence, therefore, the rays form two images of which only one is deflected from the normal, and on rotating the prism only the latter image rotates while the ordinary image remains stationary.

§ (13) GENERAL EQUATIONS REPRESENTING VIBRATIONS OF POLARISED LIGHT.—So far we have dealt only with polarised light in which the vibrations are restricted to rectilinear motion, and it was merely noted in passing that light may also be polarised in such a fashion that the particles describe circular or elliptical orbits. The relation of such vibrations to those of plane polarised light is best shown by an examination of the mathematical equations representing them.

The simple equation representing the displacement at any moment of a particle moving in simple harmonic motion in a straight line is

$$y=a \sin 2\pi\frac{t}{T},$$

where a is the amplitude, T the period of the vibration, and t the time that has elapsed since the particle was in its zero position. We may equally well look upon y as a periodic electrical force.

Now suppose this equation to represent the vibration of a plane polarised ray striking a doubly refracting plate. Let OX, OY (*Fig.* 11) represent the directions of vibration in the plate, and let OK, the direction of vibration of the incident ray, make an angle i with the direction OY. On entering the plate the vibration OK is decomposed into components along OX and OY, represented by

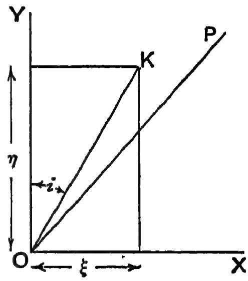


FIG. 11.

$$\xi=a \sin 2\pi\frac{t}{T} \sin i, \quad \eta=a \sin 2\pi\frac{t}{T} \cos i.$$

These two vibrations in general travel through the crystal plate with different velocities, so that if O and E represent the distances travelled in air in the times taken by the ordinary and extraordinary waves respectively to traverse the plate, the displacements will be

$$\left.\begin{aligned}\xi&=a \sin i \sin 2\pi\left(\frac{t}{T}-\frac{O}{\lambda}\right)\\ \eta&=a \cos i \sin 2\pi\left(\frac{t}{T}-\frac{E}{\lambda}\right)\\ &=a \cos i \sin 2\pi\left(\frac{t}{T}-\frac{O}{\lambda}+\frac{O-E}{\lambda}\right)\end{aligned}\right\}, \quad . \quad (1)$$

where λ is the wave-length of the light in air.

The two waves emerge parallel and the light refracted from the plate will be resultant of two superposed vibrations of the same period and amplitude but differing in phase by an amount $(O-E)/\lambda$, which varies with the thickness of the plate.

Combining equations (1) by the elimination of t we shall obtain an expression for the resultant vibration.

Let $2\pi\left(\frac{t}{T}-\frac{O}{\lambda}\right)=\theta$ and $2\pi\left(\frac{O-E}{\lambda}\right)=\phi$.

Then equations (1) may be written

$$\frac{\xi}{a\sin i}=\sin\theta,$$

$$\frac{\eta}{a\cos i}-\sin\theta\cos\phi=\cos\theta\sin\phi.$$

Squaring both sides of the second equation and substituting for $\sin^2\theta$ and $\cos^2\theta$ from the first, we obtain

$$\frac{\eta^2}{a^2\cos^2 i}+\frac{\xi^2}{a^2\sin^2 i}\cos^2\phi-2\frac{\xi\eta}{a^2\sin i\cos i}\cos\phi$$

$$=\left(1-\frac{\xi^2}{a^2\sin^2 i}\right)\sin^2\phi.$$

Hence $\frac{\xi^2}{a^2\sin^2 i}+\frac{\eta^2}{a^2\cos^2 i}-\frac{2\xi\eta\cos\phi}{a^2\sin i\cos i}=\sin^2\phi,$

or

$$\xi^2\cos^2 i+\eta^2\sin^2 i-2\xi\eta\sin i\cos i\cos 2\pi\frac{O-E}{\lambda}$$

$$=a^2\sin^2 i\cos^2 i\sin^2 2\pi\frac{O-E}{\lambda}. \quad (2)$$

This indicates that the general form of the resultant vibration is an ellipse one of whose axes is inclined to the vibration direction in the plate at an angle i.

Special cases may be considered.

(i.) *Plane Polarised Light.*—If the difference in optical path $O-E=n\lambda/2$, so that $2\pi(O-E)/\lambda=n\pi$, equation (2) reduces to

$$\xi^2\cos^2 i+\eta^2\sin^2 i\pm2\xi\eta\sin i\cos i=0,$$

or $$\xi\cos i=\pm\eta\sin i. \quad (3)$$

Hence if the plate is of such a thickness that the phase difference on emergence is equal to any whole number of half wave-lengths, the resultant vibrations are plane polarised. If n is even, $\xi/\eta=+\tan i$, and the vibrations are parallel to those of the incident light. If n is odd, $\xi/\eta=-\tan i$, and the vibrations are in a direction equally inclined to the Y axis but on the side of it remote from the incident vibration direction.

Moreover, this is the only condition which reduces equation (2) to the rectilinear form.

(ii.) *Circularly Polarised Light.*—If, however, $O-E=(2n+1)\lambda/4$, and $i=45°$, we have

$$\sin i=\cos i=1/\sqrt{2},$$

and $$2\pi(O-E)/\lambda=(2n+1)\pi/2.$$

Equation (2) then becomes

$$\xi^2+\eta^2=a^2/2, \quad (4)$$

representing uniform motion in a circle of radius equal to the amplitude of the original vibration. When, therefore, the phase difference between the two rays is equal to an odd number of quarter wave-lengths, and when the plane of polarisation of the incident ray is equally inclined to those of the rays traversing the crystal plate, the emergent ray will be circularly polarised. A crystal plate which fulfils this condition of phase difference between its emergent vibrations is known as a *quarter-wave plate* ($\lambda/4$ plate), and provides a ready method of producing circularly polarised light.

(iii.) *Elliptically Polarised Light.* — If $O-E=(2n+1)\lambda/4$, but i is not equal to 45°, equation (2) becomes

$$\frac{\xi^2}{\sin^2 i}+\frac{\eta^2}{\cos^2 i}=a^2. \quad (5)$$

That is to say, if a quarter-wave plate is so orientated that its vibration directions OX and OY are not equally inclined to the vibration direction, OK, of the incident light, the emergent ray is elliptically polarised, the axes of the ellipse being parallel to OX and OY. It is obvious that in the limiting cases when $i=0°$ or 90°, the ellipse becomes a straight line parallel to OY or OX, that is to say, the crystal transmits one ray only.

§ (14) Circularly Polarised Light. (i.) *Production by Refraction.*—Circularly polarised light can be produced, as mentioned above, by the use of a quarter-wave plate. Such a plate is most easily prepared by splitting mica along its cleavage planes into very thin sheets; by trial a sheet can then be chosen such that it produces between the transmitted rays a phase difference of $\lambda/4$ for any fixed value of λ. For use with sodium light the thickness of a quarter-wave plate of mica is 0·032 mm. It is useful not only to know the vibration directions in the plate, but also to be able to distinguish between the directions of slower and faster transmissions.

The vibration directions are easily found by holding the plate between crossed Nicols. When it is in such a position that the field remains dark its vibration directions must be parallel to those of the Nicol prisms. A method of distinguishing definitely between the directions of the fast and slow transmissions is given by R. W. Wood.[1] If the mirrors of a Michelson interferometer are adjusted to give a system of fringes in white light plane polarised in a horizontal plane, the introduction into the path of one ray of a quarter-wave plate with its vibration directions

[1] *Physical Optics*, 1914, p. 320.

horizontal and vertical will produce a retardation in that ray relative to the other, and a shift of the fringes will ensue. Now if the plate is turned through 90°, there will be a further shift in the fringes. If it is in the same direction as before, it is clear that a further retardation of that ray has taken place; that is to say, the direction which was at first horizontal is the direction of vibration of the faster moving component.

(ii.) *Production by Reflection.*—We have seen that when a plane polarised ray strikes the dividing surface between two media at the polarising angle the reflected ray and the refracted ray are polarised in perpendicular planes. Now if the conditions be those required for total reflection, both rays will be reflected along the same path, but there will be a phase difference of $\lambda/8$ between their vibrations. Two total reflections of the rays will therefore produce the phase difference of $\lambda/4$ necessary for the production of circularly polarised light, and the only other necessary condition is that the original plane of polarisation should be at 45° to the plane of incidence. Fresnel constructed a glass rhomb through which a plane polarised ray could pass (*Fig.* 12), suffering two total internal reflections

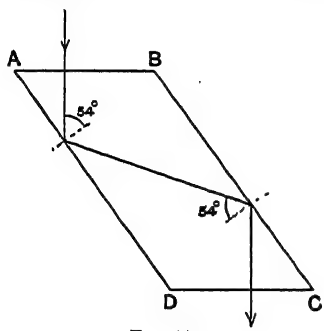


FIG. 12.

at the polarising angle of 54°, and thus attaining a phase difference of $\lambda/4$ between the two emerging components. Since the polarising angle for a glass-air surface varies very little with change of wave-length, the phase difference of the emergent rays is nearly independent of the colour of the light, and in this respect this method of production of circularly polarised light is preferable to the use of a quarter-wave plate.

(iii.) *Methods of Detection.*—It is obvious that circularly polarised light differs from both plane and elliptically polarised light by possessing no unique pair of perpendicular planes about which the path of a vibrating particle is symmetrical. Consequently when examined by an analysing prism a circularly polarised beam shows no change on rotation of the prism. In this respect it resembles unpolarised light, but can be distinguished from the latter by its behaviour in conjunction with a quarter-wave plate, for if the plate be so orientated that it neutralises the phase difference existing between the two components of the circularly polarised light the emergent light will be converted to plane polarised light, which can then be extinguished by a suitable orientation of the analysing prism.

It is of use in some connections to determine the direction of rotation of circular vibrations, which, if produced by a crystal plate, will vary according to the character of the crystal and the plane of polarisation of the incident light. This will be clear on reference to *Fig.* 11. Suppose OP represents the vibration direction of the incident light. The force moving a particle from O towards P will be resolved in the plate into two perpendicular forces tending to move the particle from O towards X and Y, where OX and OY are the vibration directions of the ordinary and extraordinary rays respectively. In a positive crystal the ordinary ray travels the faster, so that the vibration along OX, say, is executed more quickly than that along OY. After a certain interval of time there is a phase difference between the rays of $\lambda/4$, that is to say, while the particle is about to move from X towards O there is an equal force propelling it in a perpendicular direction upwards. Hence it will move by a circular path in a counter-clockwise direction from X to Y. If at this moment the rays emerge from the plate there will be no further change in the phase difference, and the vibration will continue to be circular and counter-clockwise. In a negative crystal it is obvious that under the same conditions the path of a vibrating particle would be clockwise. These deductions, together with a consideration of the case when the vibration direction of the incident ray is perpendicular to OP, leads to the rule that if the vibration direction of the incident light is turned in a clockwise direction through 45° from the direction of fastest vibration in the crystal plate, the circular path of the emerging vibrations will be described in a clockwise direction also.

The direction of vibration can be found experimentally by means of a quarter-wave plate for which the direction of the faster vibration is known. By a suitable orientation of the plate the phase difference of the light under examination can be neutralised or doubled, with the production of light plane polarised parallel or perpendicular to the incident light, and the directions of fast and slow vibrations in the unknown plate can be deduced. An application of the construction given above then determines the rotational direction.

§ (15) Elliptically Polarised Light. (i.) *Production.*—From what has already been said it is evident that the conditions under which circularly polarised light is produced are special cases of the conditions necessary to produce elliptically polarised light. The production of the latter, therefore, generally requires the fulfilment of one condition less than the number necessary for the production of circular polarisation. For example, dealing first with methods of production by refraction, if a quarter-wave plate is used, elliptically polarised light is produced if the plane of polarisation makes any angle with the vibrations in the plate between 0° and 90°; at an angle of 45° the elliptical vibration becomes circular. The condition regulating the thickness of the plate for the production of circular polarisation may be relaxed for elliptical polarisation, which will result from the passage of plane polarised light through a thin plate producing any phase difference that is not a multiple of $\lambda/4$, but in this case the axes of the ellipse will be inclined to the original direction of vibration.

For production by a reflection method, it is obvious that the necessary conditions are fulfilled by a single internal total reflection in a glass prism of plane polarised ray at the polarising angle, for this produces a phase difference of $\lambda/8$ between two coincident perpendicularly polarised rays. When the plane of polarisation of the incident light is inclined to the plane of incidence at an angle of 45° the light reflected from any medium is to some extent elliptically polarised. With transparent media the eccentricity of the ellipse is so great that the vibrations are very nearly linear, but if the reflection takes place from a metallic surface the ellipticity of the polarisation is pronounced, and when the incident light is polarised at an angle of 45° with the angle of incidence the polarisation of the reflected light is nearly circular.

Drude[1] explains this production of elliptical polarisation by postulating a gradual, in place of an abrupt, change of refractive index at the surface of the medium. Further work by Lummer and Sorge[2] on solid media and by Lord Rayleigh[3] on liquid surfaces shows that a higher surface refractive index, due in the former case to the action of polishing materials and in the latter to the presence of a thin film of grease, tends to produce elliptical polarisation of reflection, but the two former authors have also shown that in some cases the phenomenon is probably due to the presence of surface strains in the medium, and is affected by subjecting the medium to pressure.

(ii.) *Methods of Detection and Analysis.*—When examined with an analysing prism elliptically polarised light shows a waxing and waning of intensity on rotation of the prism, the maximum and minimum points occurring when the vibration direction of the Nicol is parallel to the longer and shorter axes of the ellipse respectively, but no orientation of the analyser wholly extinguishes the light. In this respect elliptically polarised light behaves like a mixture of plane polarised and unpolarised light, but it can be distinguished from the latter, as in the similar case of circularly polarised light, by its reduction to plane polarised light by means of a quarter-wave plate suitably orientated. The direction of the axes of ellipse can be deduced at the same time, for they must of necessity be parallel to the principal planes of the quarter-wave plate when it is in position to render the emergent light plane polarised. The ratio of the axes can be calculated by determining, by means of a Nicol prism, the angle between the vibration direction of the emergent plane polarised light and the directions of the axes of the ellipse.

(iii.) *Babinet's Compensator.*—A more accurate method of determining the constants of an elliptical vibration is by the use of *Babinet's Compensator.* This consists of two rectangular prisms of quartz (*Fig.* 13) placed with their hypotenuses in contact so that together they form a plane parallel plate, the thickness of which can be varied by sliding the prisms along their interface. The optic axes are parallel and perpendicular to the plane of the paper in the two halves as represented by the shading of each prism. Plane polarised monochromatic light entering the plate normally is resolved into two rays whose separation is further increased on entering the second prism. The relative retardation produced between the rays by their passage through the whole plate is proportional to $(e_1 - e_2)(\mu_e - \mu_o)$, where e_1 and e_2 are the distances traversed in the two prisms, and μ_e and μ_o are the two extreme refractive indices. Along certain lines parallel to the prismatic edges the phase difference between the emergent rays will be equal to a whole number of wave-lengths; there will therefore be equidistant lines of plane polarised light occurring in the emergent light, which can be made to appear as dark bands by the use of an analysing prism crossed with reference to the polariser. Between the bands the light will be elliptically polarised in general; half-way between each band the light will be plane polarised in a direction oblique to the plane of polarisation of the first set, being perpendicular to it. If the plane of polarisation of the incident light

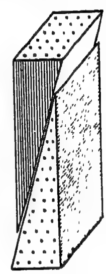


Fig. 13.

[1] *Lehrbuch der Optik*, 1900, p. 266, Engl. trans., 1920, p. 287.
[2] *Ann. d. Physik*, 1910, xxi. 325.
[3] *Phil. Mag.*, 1908, xvi. 444; *Scientific Papers*, iii. 496.

is at 45° to the vibration directions of the prisms, this condition gives the best definition to the bands when examined with an analyser.

The compensator can be used as a measure of phase difference, since the space between each dark band or graduation corresponds to a difference in retardation of $\lambda/2$. If now one prism is slid over the other through a certain distance, the system of dark bands moves unchanged through half that distance. If a is the linear distance between two adjacent bands, a shift of the prisms by an amount x will produce a change in the retardation at any point of $\frac{1}{2}\lambda x/a$. Any change taking place in the state of polarisation of the transmitted light will be accompanied by a change in the retardation produced, and this can be measured by the relative alteration in position of a dark band. The original form of compensator was provided with a cross-wire controlled by a micrometer screw, the prisms being fixed in position; Jamin adapted the instrument to the special study of light vibrations, and replaced the movable cross-wire by a fixed wire, at the same time connecting one prism to the micrometer screw.

To determine the phase difference of the two components forming an elliptically polarised ray the cross-wire of the instrument is set on the central dark band, corresponding to zero phase difference. Elliptically polarised light is then substituted for the incident plane polarised light, and the quartz wedge is moved by the micrometer screw to bring the central band and cross-wire into coincidence again. This gives a measure of the phase difference producing the elliptical vibrations; if the compensator is rotated until the phase difference is $\lambda/4$, the axes of the elliptical vibration will be parallel to the vibration directions in the quartz prisms, while their ratio is given by the tangent of the angle between one of these directions and the vibration direction of the analysing Nicol.

(iv.) *Savart's Polariscope.*—*Savart's polariscope*, which is used for the detection of small quantities of plane polarised light in a partially polarised beam, is constructed to give similar parallel dark bands. A plane parallel plate is cut from a quartz crystal at an angle of 45° to the optic axis. The plate is then sectioned parallel to the plane surfaces, one half is rotated through 90° relative to the other, and the two are cemented together. The greatest sensitivity is attained when the direction of the polarised vibration is perpendicular to the dark bands produced between crossed Nicols; in this position as little as 1 per cent of polarised light can be detected.

§ (16) Behaviour of Crystal Sections in Polarised Light.—One of the chief applications of polarised light is to the examination of thin crystal sections, for the variation in their behaviour towards transmitted light is one of the most valuable criteria for the identification of crystals.[1] In general the crystal to be examined is obtained in the form of a plane-parallel plate, too thin to cause the total separation of the ordinary and extraordinary rays when both are produced, but thick enough to give clear indications of the isotropic or birefringent nature of the crystal and its general optical character.

The most essential instrument for the examination of crystal sections is some form of petrographical microscope, fitted with removable polarising and analysing Nicols which can be rotated together relative to the specimen or separately relative to one another. The microscope should also possess fittings to hold a quarter-wave plate, quartz wedge or gypsum plate, the use of which will appear later. The whole subject of crystal examination is vast and intricate; it is proposed to give here only a brief indication of the main characteristics of crystal sections and the general methods employed. It must be borne in mind that, although it is mainly the characteristics of principal sections of crystals that are dealt with, an examination of a microscopic slide will in general show crystal sections cut in all directions, in which case their properties will be modifications of those described.

There are two main classes of phenomena displayed by crystal sections: firstly, those displayed in parallel plane-polarised light, and, secondly, those occurring when the polarised rays are convergent. The former class, as being the simpler, will be considered first.

§ (17) Crystal Sections in Parallel Light.—It has been shown that in many respects the character of a crystal may vary in different directions. This may be applicable also to its powers of absorbing light.

(i.) *Pleochroism.*—Some crystals show different powers of selective absorption for the ordinary and extraordinary rays, with the result that a section parallel to the axis may appear of one colour when the ordinary ray only is transmitted and a different shade of the same colour, or an entirely different colour, when it is seen by extraordinary rays only. The coloured mica known as biotite is a good example of a *pleochroic crystal*, as it is termed; the extreme case of tourmaline has already been mentioned, and in sections not sufficiently thick to absorb the ordinary ray completely there is a variation in colour from a pale bluish-green to an almost opaque brown as the plane of polarisation of the incident light is rotated.

(ii.) *Interference Tints.*—The appearance of crystal plates between a pair of Nicols depends

[1] See "Crystallography," § (21).

on the nature of the rays emerging from the crystal. In general the two rays produced by a birefringent crystal emerge with a definite phase difference depending upon the nature of the crystal, its thickness, and the wave-length of light used. Assuming that the section is of a standard thickness, so that the effects produced by different sections can be correlated easily,[1] the polarisation of the emergent light will depend upon the wave-length of the light; if white light is used, the difference in retardation for different wave-lengths will lead to the plane polarisation of light of some colours and elliptical or circular polarisation of light of other colours. If the Nicols are crossed, the observer will lose light of all wave-lengths for which the phase difference in the plate is any multiple of λ (remembering that the resolution of the two perpendicularly polarised rays to the vibration direction of the analyser causes a further retardation of $\lambda/2$), and the resultant light will possess the colour complementary to that absorbed. The colours corresponding to retardations of varying amounts are given in the well-known Newton Scale of Interference Colours. A crystal of low birefringence will produce a small phase difference equivalent only to one wave-length for some particular shade. That colour alone will be eliminated, and a tint of the first order appears, clear grey, yellow, or red; higher birefringence entails the loss of more than one colour, and is indicated by the brilliant second order tints, while very powerful double refraction gives the soft impure tints of still higher orders.

(iii.) *Extinction.*—If the vibration direction of the analyser is parallel to one of the vibration directions in the section, the polarised light will pass through the crystal unresolved, and will be cut off completely by the analyser. Thus if a crystal section is revolved in its own plane between crossed Nicols, in four perpendicular positions the field appears dark; these are known as the *positions of extinction*, and the line in the section then parallel to the vibration direction of the polariser is known as the *direction of extinction*. The angle between the direction of extinction and a crystallographic axis indicates the nature of the crystal; uniaxial crystals have "straight extinction," that is to say, their extinction angle is zero. Accurate measurement of this angle, either with the microscope or by other methods, is of great importance in the systematic investigation of a crystal. Sections of birefringent crystals perpendicular to an optic axis and all sections of a cubic crystal will, of course, appear dark in all positions between crossed Nicols.

(iv.) *Sign of Birefringence.*—The sign of the birefringence may be deduced by the use of a *gypsum plate*. Such a plate is a thin cleavage section of the monoclinic crystal gypsum (selenite), the cleavage being parallel to the plane of symmetry, which contains the optic axes. A section of suitable thickness shows a uniform interference tint of first order red between crossed Nicols, and a slight change in the phase difference is enough to change the tint sharply to the lower first order blue-grey or the higher second order yellow. Now, if a crystal section whose extinction directions are known is superposed on the gypsum plate between crossed Nicols, so that one extinction direction, that is to say, the vibration direction of one ray, is parallel to one vibration direction in the plate, a raising of the colour tint to yellow will indicate that the directions of fast vibration in the two crystals coincide. It is only necessary to have the directions of fast and slow vibrations marked on the gypsum plate to be able to deduce at once the corresponding directions in the crystal, and thence by reference to the position of its optic axis, or axes, to determine the sign of its birefringence.

§ (18) Crystal Sections in Convergent Light.—The second class of phenomena shown in the examination of crystal plates is produced when the section is viewed in convergent or divergent polarised light, and takes the form of the well-known "rings and brushes" produced by interference. For this purpose the microscope has practically to be converted into a telescope; details of the necessary arrangement may be found in Spitta's *Microscopy*, p. 203 (1907), or in F. E. Wright's *Methods of Petrographic Microscopic Research*, p. 39 (1911). (To the latter work the reader is also referred for a very detailed account of the apparatus and methods used in the exact microscopic measurement of the optical constants of a crystal.) With this arrangement any point on the field of view corresponds not to any special point on the section, but gives the total effect of all rays travelling through the section at a certain angle, the centre of the field corresponding, of course, to the effect produced by light parallel to the axis of the system and normal to the section.

(i.) *Uniaxial Crystals.*—Let us consider first the behaviour of a plate cut from a uniaxial crystal perpendicular to the axis. Suppose the light to be diverging from S, and polarised to vibrate vertically (*Fig.* 14A), and let P represent the crystal section whose optic axis is in the direction of OS. Any oblique ray will emerge from the plate with a definite phase difference between its two perpendicularly polarised components. The phase difference will vary with the angle of obliquity, assuming that the thickness of the plate and the wave-length are constant. Hence

[1] The thickness of a rock section or crystal section cut for general examination is usually between 0·01 mm. and 0·02 mm., producing in quartz the clear grey and yellow of the first order interference tints.

for a certain ray Sp the phase difference produced by the plate will have a value λ, and when examined through an analyser the corresponding point in the field of view will appear dark. The locus of all points corresponding to rays passing through the plate with the same obliquity but varying azimuth, must by symmetry be a circle round O; consequently there will be a dark circular

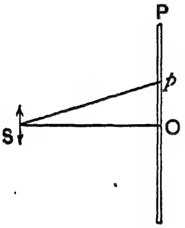


FIG. 14A.

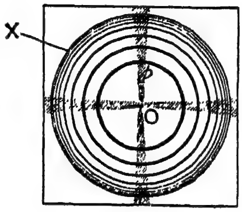


FIG. 14B.

fringe visible marking a phase difference of λ. Similarly, there will be other concentric circles marking the locus of points where the retardation is 2λ, 3λ, . . . $n\lambda$, their separation decreasing towards the edges of the field. Moreover, any ray lying in the planes parallel or perpendicular to the vibration direction of the polariser will not undergo resolution in passing through the plane, and will be cut out

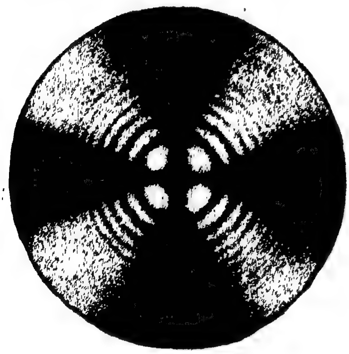


FIG. 16.

FIG. 17.

entirely by the analyser. This will result in the field of view being crossed by dark rectangular diagonals parallel to the vibration directions of the polariser and analyser, and the appearance of the field will be similar to that shown in *Fig.* 14B. With monochromatic light the fringes will be black; white light will give coloured fringes. The black fringes, known as *isochromatic lines*, represent the intersection with the crystal plate of the *isochromatic surfaces*, which are the loci in space of all points at which the phase difference is the same. For uniaxial crystals each isochromatic surface is formed by the revolution of a hyperbolic curve about the optic axis. This surface is shown in *Fig.* 15, together with the more complicated form for biaxial crystals. Sections of a uniaxial crystal perpendicular to the optic axis show circular fringes, as we have already

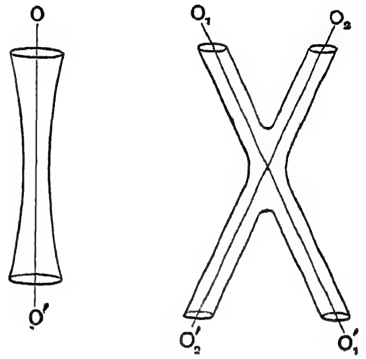


FIG. 15.

seen; in sections parallel to the optic axis the intersection with the isochromatic surfaces will be hyperbolae. Photographs of the interference figures for uniaxial crystals are shown in *Figs.* 16 and 17. The case of quartz and other optically active crystals is worthy of comment. The interference figure for quartz cut perpendicular to the axis is shown in *Fig.* 18. It will be noted that it resembles that of other uniaxial crystals except in that part representing rays travelling along or close to the optic axis. As will be shown later, the rotatory power of quartz leads to the deduction that a ray parallel to the axis is resolved into two circularly polarised rays which combine on emergence to form light polarised in a direction dependent on the thickness of the crystal. Except for certain thicknesses of quartz, therefore, the centre

of the interference pattern produced will show bright between crossed Nicols.

(ii.) *Biaxial Crystals.*—In the case of biaxial crystals a section perpendicular to the acute bisectrix will show the isochromatic lines as a family of lemniscates enclosing the points corresponding to light travelling along the optic axes. These will be crossed by dark brushes, the position of which varies with the

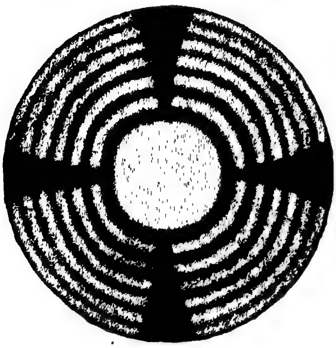


FIG. 18.

position of the optic axial plane in the section relative to the vibration directions of the Nicols. *Fig.* 19 shows the appearance of the fringes when these directions are parallel;

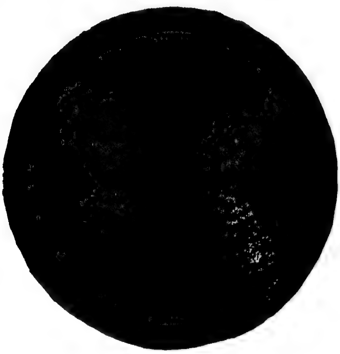


FIG. 19.

if the section is rotated through 45° the brushes become hyperbolae, whose vertices are at the points corresponding to the directions of the optic axes (*Fig.* 20). If white light is used the effect is complicated by dispersion of the optic axes and bisectrices.

(iii.) *Measurement of Optic Axial Angle.*—It is this latter 45° position that is used for measuring the optic axial angle. By tilting the section about an axis in its plane perpendicular to the line joining the optic axial points, the position of the hyperbolic fringes is changed, and the tilt can be adjusted until one vertex lies on the cross wires of the microscope. Then rays parallel to the axis of the microscope must travel along an optic axis. If both vertices

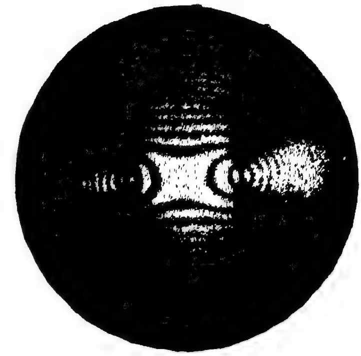


FIG. 20.

are brought in turn on to the cross wires the angle of tilt between the two positions gives the angle between the optic axes in air. It may be necessary, if the optic axial angle is very large or very small, to use a liquid of high refractive index between the objective and the section, so that both axial rays can be brought normal to the field in turn; by obtaining measurements in sections cut perpendicularly to the acute and obtuse bisectrices the actual optic axial angle in the crystal can be obtained.[1]

(iv.) *Sign of Birefringence.*—Mention must be also made of the use of a quarter-wave plate and a quartz wedge in determining from the interference figure the sign of birefringence of a crystal. Suppose a quarter-wave plate of mica is inserted between the analyser

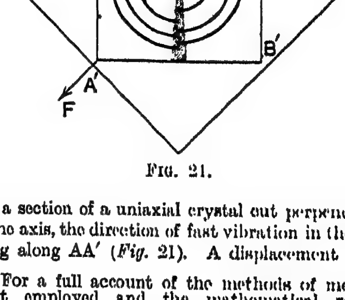


FIG. 21.

and a section of a uniaxial crystal cut perpendicular to the axis, the direction of fast vibration in the mica being along AA′ (*Fig.* 21). A displacement of the

[1] For a full account of the methods of measurement employed and the mathematical relation between the apparent and true optic axial angles see F. E. Wright (*loc. cit.*), and *Manual of Petrographic Methods*, Johannsen, p. 102.

ring system will occur, adjacent quadrants expanding and contracting. If the crystal is positive the rings in quadrants A and A′ will expand and those in B and B′ will contract; with a negative crystal the result will be reversed. This follows very simply from a consideration of the retardation at a point P, say, in the quadrant B. If P lies on BB′ the retardation in the crystal will take place between rays vibrating parallel to OB and OA, and if the crystal is positive, OA will be the direction of faster vibration, which therefore coincides with the faster vibration direction of the mica plate. The retardations of the crystal and the mica will therefore be additive, and the total phase difference increased; this will apply to all points in the quadrants B and B′. Hence a smaller thickness of the crystal will, together with the mica plate, produce a retardation of λ, and consequently in these quadrants the fringes will close in towards the centre. In the alternate quadrants A and A′ a corresponding decrease in the total phase difference will produce a movement of the rings away from the centre.

With a biaxial crystal a similar effect is observed, and the closing of the curves in the quadrants joined by the direction of slow vibration in the mica plate again denotes a positive birefringence. The movement is not always easy to trace, however, and a better indication is obtained by the use of an elongated *quartz wedge* cut with its line of greatest slope and its edges parallel to the optic axis. The section is placed with the trace of its optic axial plane at 45° to the vibration direction of the incident light, and the thin end of the wedge is inserted along the line of the optic axes. The faster vibration in the quartz is perpendicular to the long axis; if the crystal is positive its direction of faster vibration is perpendicular to the axial plane; hence the quartz and the crystal will have a subtractive effect, and the fringes will open as the wedge is pushed further in and a greater thickness of quartz is brought into operation. With a positive crystal there is a corresponding contraction of the fringes as the wedge is inserted parallel to the line joining the optic axes.

§ (19) Detection and Measurement of Strain in Isotropic Materials.—An important application of polarised light is its use in testing glass for strain. Brewster called attention to the production of birefringence in isotropic media by mechanical stresses, and later Clerk Maxwell [1] attacked the problem of a strained plate mathematically. A full description of the methods used will be found elsewhere and will not be treated of further here.

An analogous method is also employed to obtain quantitative measurements of stresses present in mechanical constructions, although at present the whole subject has hardly passed beyond the experimental stage. A comprehensive summary is given by Low in the *Aeronautical Journal*, 1918, xxii., together with a full bibliography of other researches and publications.[2] The general method may be outlined here.

A model of a mechanical construction is cut out of an isotropic medium; glass has been used, but xylonite, although not so transparent, serves the purpose better, inasmuch as it is easier to work and can be obtained in thick sheets showing no strain, and resembles in its mechanical properties the actual metals used in the full-scale construction. The presence of stresses in the model causes the stressed portion to act as a birefringent medium, and transmitted light is resolved into two rays polarised parallel and perpendicular to the direction of stress. The rays travel with different speeds, and as in the case of a thin crystal plate, the emergent light is coloured owing to the elimination of certain wave-lengths by interference. The tint produced varies with the retardation between the rays, which in turn is dependent on the amount of stress present. It is possible, therefore, by means of a block subjected to known stresses, to tabulate for any given thickness of the medium the stresses corresponding to the various tints produced, and from this table to deduce the stress present in any model giving a certain interference tint. A more accurate method of quantitative measurement depends upon the fact that while two superposed similar stresses have a simple additive effect if parallel, the total effect of two similar perpendicular stresses is proportional to their difference. Thus an unknown stress coloration may be balanced by superposing a plate subjected to a stress of known amount and direction to produce between crossed Nicols a uniformly dark field.

It is of importance to note that the birefringence is dependent upon the stress present and not upon the strain; a plate strained beyond its elastic limit may show no optical effect if the actual stress is removed.

§ (20) Rotatory Polarisation (i.) *Crystals.*—When discussing the behaviour of crystals in the transmission of polarised light, the optic axis was defined as a direction in the crystal along which light travels without division of path. While in many crystals this is equivalent to a statement that polarised light travelling along an optic axis emerges from the crystal unchanged, some crystals, most notably quartz, have the power in the direction of the axis of rotating the plane of polarisation. All such crystals, known as *optically active substances*, belong to those classes of their respective crystallographic systems which possess no plane of symmetry; they are all enantiomorphous, all but one class possessing only symmetry of rotation about an axis, the other class possessing no degree of symmetry at all. With these crystals two forms are possible, each being the mirror image of the other, but not capable of complete identification with the other by any change of orienta-

[1] *Roy. Soc. Edin. Trans.* xx. Part I.; *Scientific Papers*, i. 30.
[2] See also *Roy. Inst. Proc.*, 1916, xiv., Lowry.

tion. In all cases in which the two forms have been obtained it is found that similar plates of the two cut perpendicular to the optic axis rotate the plane of polarisation of transmitted light by equal amounts but in opposite directions. The two crystal forms of quartz are well known, and can be distinguished morphologically by the position relative to the fundamental prismatic and pyramidal faces of six pairs of small faces occurring three at each end of the crystal and known in the Millerian system of notation as $\{41\bar{2}\}$ and $\{4\bar{1}\bar{2}\}$. The second of the pair is often absent, but the presence on the $\{41\bar{2}\}$ face of striae parallel to its line of intersection with the $\{4\bar{1}\bar{2}\}$ face indicates what would be the position of the latter if developed, and shows the right- or left-handed nature of the crystal.

There is considerable confusion[1] in the definitions given by various authors of "right-handed" and "left-handed" rotation of the vibration plane; this, no doubt, is due partly to the fact that the convention adopted is based on the aspect of the phenomenon as it appears to an observer, and is not descriptive of the motion taking place along the path of the ray itself. An optically active substance is said to produce *right-handed rotation*, or to be *dextro-rotatory*, when to an observer looking along the path of the light towards its source the vibration direction is rotated in a clockwise direction; if the rotation is counter-clockwise the substance is said to be *laevo-rotatory*.

Examples of crystals possessing rotatory power are found in the various crystal systems; sodium chlorate and sodium bromate, crystallising in the cubic system, and therefore possessing no unique axis, show optical activity to the same extent in all directions; sulphate of strychnia, a tetragonal crystal, and the double sulphate of potassium and lithium which belongs to the hexagonal system are other examples, while cinnabar, a trigonal crystal, shows a rotatory power nearly twenty times as strong as that of quartz.

(ii.). *Liquids and Gases.*—Many optically active crystals lose their power when in solution, or when in a non-crystalline state; quartz in solution in potash or in a fused amorphous state is inactive, but sulphate of strychnia is an exception to this rule, and shows activity both when crystalline and when dissolved. This power is shared by many organic liquids and solutions of organic compounds; in the process of other investigations Biot accidentally discovered the laevo-activity of oil of turpentine, and on inquiring further into the matter found that a similar activity was displayed by many other liquids. The rotatory power of sugar solutions is well known, and forms the basis of the universal method of testing and examining sugars. [A separate article dealing with polarimetry and saccharimetry will be found elsewhere, to which reference may be made for full information on the subject.] Substances which are optically active in solution were found by Biot to retain their power in the solid state if they did not crystallise; since all these on solidification formed biaxial crystals, the signs of rotatory power were completely masked by birefringent effects. Later work by Pocklington[2] and Dufet[3] on cane-sugar and various salts crystallising in the rhombic and monoclinic systems has demonstrated the presence of optical activity in the crystals and the possibility of measuring its value along an optic axis. Biot further showed that the transformation of a substance to a gaseous state does not destroy its optical activity, and later Gernez[4] proved by experiment that the rotatory power of a certain tube-length of vapour is equal to that of the column of equal cross-section of liquid into which it condenses.

§ (21) Optically Active Substances. (i.) *Rotatory Power.*—Apparently, then, there are two classes of optically active substances; some substances, like quartz, depend for their power on their crystalline state, that is, presumably, on the grouping of the molecules forming their crystals, while in others the rotatory power appears to be an inherent property of each molecule, since their separation by solution or vaporisation does not affect the power. This conception led to the introduction of the term "*molecular rotatory power*"; if a solution of density d containing p grams of an optically active substance in q grams of solvent is contained in a tube l decimetres long, and if R is the angular rotation produced by the tube,

$$\mathrm{R}=(\rho)\frac{pd}{p+q}\cdot l,$$

$pd/(p+q)$ being the amount by weight of the substance in unit volume of the liquid. (ρ) was taken to be a constant for the substance for any specified temperature and wave-length, and was denoted by the *molecular rotatory power* or *specific rotation* of the substance. Putting $q=0$ and $l=1$, $(\rho)=\mathrm{R}/d$; hence the molecular rotatory power may be more specifically defined as the angular rotation of 1 decimetre of the substance in the pure solid state divided by its density. This is the definition which should be adopted, for Biot showed that the angular rotation produced is not strictly proportional to the amount of substance

[1] See "Polarimetry," § (1); "Quartz, Optical Rotatory Power of," § (1).

[2] *Phil. Mag.*, 1901, ii. 368.

[3] *Jour. de Phys.*, 1904, iii. 757; *Bull. Soc. Fran. de Min.*, 1904, xxvii. 156.

[4] *Soc. Phys. Séances* and *Comptes Rendus*—many publications between 1887 and 1892.

present. Landolt,[1] after careful investigation, deduced the formula

$$R = A + Bq + Cq^2,$$

where R is the rotation produced by 10 cm. of a solution containing q parts by weight of the solvent in 100 parts by weight of the solution, A is the molecular rotatory power of the pure substance, and B and C are constants to be determined for each substance and solvent by observations on solutions of different strengths. It should be possible, if B or C is negative, for the same solution to exhibit positive and negative rotation at different concentrations, and this has been shown to occur in the case of malic acid, which is dextro-rotatory in concentrated solutions, and laevo-rotatory when dilute, a solution containing 65·7 per cent of water being inactive.

The effect of temperature on the angle of rotation is considerable, and is of importance, especially in connection with polarimetric work; it will be treated under that head, therefore, and not further described here.

(ii.) *Variation with Wave-length of Light.*—The variation of the rotation with wave-length is, however, of more far-reaching importance. Biot's observations led to the result that the angle of rotation is approximately inversely proportional to the square of the wave-length. From Biot's figures for quartz Stephan proposed the empirical formula

$$R = -1{\cdot}581 + \frac{0{\cdot}80403}{\lambda^2},$$

which gives results in fair accordance with the observed values, the first term representing the amount of departure from the inverse square law. Boltzmann[2] suggested the addition of a fourth order term to Biot's simple proportion formula, putting

$$R = \frac{k_1}{\lambda^2} + \frac{k_2}{\lambda^4} + \ldots,$$

which yielded more accurate results and was applicable to substances other than quartz.

More modern investigations of the problem have led to the development of equations of the form

$$R = k + \frac{k_1}{\lambda^2 - \lambda_1^2} + \frac{k_2}{\lambda^2 - \lambda_2^2} + \ldots,$$

where k, k_1, k_2, . . . λ_1, λ_2, . . . are constants on which the properties of the medium depend. This expression with the five constants indicated has been verified to a high degree of accuracy by Lowry.[3]

[1] *Berichte der Deutschen Chemischen Gesellschaft*, 1880, xiii. 2329.
[2] *Pogg. Annal.*, 1874, 128. See also Peddie, *Roy. Soc. Edinburgh Proc.*, 1882, xi. 815.
[3] *Phil. Trans. Roy. Soc.*, 1912. A, ccxii. 261. See also "Quartz, Optical Rotatory Power of."

§ (22) Applications. (i.) *General.*—This variation of rotation with wave-length affords a simple means of distinguishing between right- and left-handed rotatory power. If plane polarised white light is transmitted through an optically active substance, any orientation of the analyser can only eliminate light of one wave-length, so that the final beam may show any tint contained in the Newton scale of interference tints. Since violet light suffers greater rotation than the rays of longer wave-length, the colour produced by a dextro-rotatory substance will change from the shade complementary to red to the shade complementary to violet as the analyser is given a clockwise rotation. The sequence will be reversed if the substance is laevo-rotatory, and a clockwise rotation of the analyser will produce a change from a reddish tint through the so-called "transition violet" to bluish green. The transition violet marks the point at which the most intense part of the spectrum, the yellow region, is eliminated; a very slight rotation of the analyser is sufficient to cause a marked change of colour at this point, to red on the one side and to blue on the other.

(ii.) *Biquartz.*—This sudden change has been utilised in the *Biquartz plate*, which is of great use in setting the planes of analyser and polariser accurately perpendicular, or in detecting very small amounts of rotation of the vibration plane. The biquartz consists of two equal plane-parallel semicircular plates cut perpendicular to the axis from right- and left-handed quartz crystals, and cemented together to form a circular plate. The thickness of the plate is designed to produce the transition violet tint when placed between two Nicols; if the thickness is 3·75 mm. the tint is produced between parallel Nicols, while a plate of double this thickness gives the same effect with crossed Nicols. A rotation of the vibration plane of the analyser, or of the incident light, tends to produce a change in the colour of one half of the plate towards red of the first order of Newton's scale, and in the other half to a blue or green of the second order, and the juxtaposition of the two halves makes it possible to detect a very small alteration of the vibration direction. If the biquartz is used with sodium light there will be a corresponding difference of intensity produced between the two halves, which affords an almost equally sensitive test.

A single quartz plate cut from either a right- or a left-handed crystal, and producing between crossed Nicols a uniform transition violet colour, sometimes known as a *Biot quartz plate*, may be employed instead of a gypsum plate, the use of which has been described in dealing with the microscopic

examination of double refraction in crystal sections.

(iii.) *Theoretical Considerations.*—It has been suggested that optically active crystals owe their power to the arrangement of molecules in the crystal. Further support was lent to this theory by Reusch, who successfully imitated a plate of quartz cut perpendicular to the axis by a pile of thin mica plates so arranged that their maximum velocity axes formed a spiral. The thinner the plates and the greater their number, the more complete was the imitation. Later Ewell [1] showed that a similar effect is obtained by passing light along the axis of a twisted cylinder of gelatine.

In the case of liquids, solutions, and gases, where no definite arrangement of the molecules can be postulated, it is of importance to note that in all cases the optically active substance contains the tetravalent carbon atom, and it is an obvious inference that just as similar molecules may be grouped together to form crystals related to each other as object and mirror image, so the four atoms or univalent groups of atoms linked to the central carbon atom may be arranged in two distinct ways to produce molecules similarly related. Le Bel and Van't Hoff [2] explained the occurrence of dextro- and laevo-rotatory liquids along these lines.

It is impossible here to enter into a full theoretical discussion of the phenomenon of optical rotation; a complete explanation has been advanced from the standpoint of the electromagnetic theory, and is to be found in various advanced text-books, more especially that of Drude,[3] whose treatment is followed in outline by R. W. Wood [4] (see "Quartz, Optical Rotatory Power of").

An explanation must necessarily involve a more precise knowledge of the nature of the vibrations that are transmitted along the optic axis of a rotatory crystal. The earliest information on this point, due to Fresnel, was based on the fact that circularly polarised light passes unchanged along the optic axis of a quartz crystal. A linear vibration may be regarded as the equivalent of two superposed and oppositely described circular vibrations of equal period and amplitude. Fresnel assumed that plane polarised vibrations, on entering a crystal along the optic axis, are resolved into two such circular vibrations, one of which travels with a greater velocity than the other. It is easily shown that in such a case, if the vibrations emerge from the crystal at the moment when one is retarded by half a wave-length relatively to the other, they will compound to form a linear vibration whose direction is at 90° with that of the incident light, that is to say, the crystal has caused a rotation of the plane of polarisation by 90° in the direction of rotation of the faster moving vibration.

(iv.) *Fresnel's Compound Prism.*—This simple theory explains the dependence of the angle of rotation upon the thickness of the medium traversed and upon the wave-length of light; moreover, Fresnel was able to demonstrate the real existence of two circularly polarised rays within a crystal by showing that they were capable of complete separation and could each be analysed after leaving the crystal. This he accomplished by building up a compound rectangular block of alternate prisms of right- and left-handed quartz (*Fig.* 22). At each interface the refractive indices

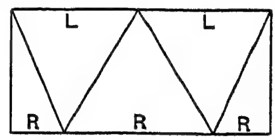


FIG. 22.

for the rays are interchanged, since the slower travelling ray in right-handed quartz becomes the faster moving ray in left-handed quartz, and the obliquity of the interfaces is so arranged as to increase the consequent separation of the rays at each refraction, until the emergent rays are completely separated. On analysis by means of a quarter-wave plate the rays are found to be circularly polarised in opposite directions.

It may be noted in passing that precautions have to be taken in the construction of quartz prisms for spectroscopic work to eliminate this separation of the rays. Cornu showed that a 60° prism will produce a separation of 27″ between the rays of sodium light travelling along the optic axis, but by using a prism composed of two similar halves of right- and left-handed quartz the separation produced in the first half is exactly neutralised by that produced in the second half, and this is the common type of prism in use at the present.

(v.) *Form of Wave-surfaces in Quartz.*—This phenomenon of double refraction along the optic axis of quartz necessitates a modification in our conception of the wave-surfaces within the crystal. The sphere and spheroid which in inactive uniaxial crystals touch at the extremities of the axis are not in similar contact in the case of quartz. Moreover, it has been shown that in addition to an increase in velocity, or a decrease in refractive index, for the extraordinary ray as its inclination to the optic axis decreases, there is a very slight increase in the velocity of the ordinary ray, only noticeable when its inclination to the

[1] *Am. Journ. Sci.*, 1890, viii. 89.
[2] *Soc. Chim. Bull.*, 1874, ii. 22, 337.
[3] *Lehrbuch der Optik*, 1900, p. 638, Engl. trans., 1920, p. 400.
[4] *Physical Optics*, R. W. Wood. 1914.

axis is very small, so that it appears that in the case of quartz the extraordinary wave-front is slightly depressed, and the ordinary wave-front elongated, in the direction of the optic axis.

(vi.) *M'Cullagh's Theory.*—A mathematical investigation of the double refraction of quartz in all directions was published by Airy in 1831.[1] As his hypothesis he assumed that for the general case both rays were elliptically polarised, with their major axes perpendicular and parallel to the principal plane of the crystal, that the rotatory vibrations were in opposite directions, and the ratio of their axes was equal, becoming unity when the rays were parallel to the optic axis and infinity when perpendicular to the axis. Five years later M'Cullagh attacked the problem again, and, starting with the ordinary differential equations representing wave-motion, introduced arbitrary modifications which made them inclusive of the effects of rotatory power.

The ordinary equations are:

$$\frac{\delta^2\xi}{\delta t^2}=b^2\frac{\delta^2\xi}{\delta z^2},\quad \frac{\delta^2\eta}{\delta t^2}=b^2\frac{\delta^2\eta}{\delta z^2},$$

where ξ, η are the displacements in a plane perpendicular to the optic axis of a particle whose ordinates are x, y, and z relative to the same axes, the optic axis coinciding with the z axis; b is the velocity of propagation of the wave-front along the axis. To these M'Cullagh added an arbitrary term, giving the equations the form

$$\frac{\delta^2\xi^2}{\delta t^2}=b^2\frac{\delta^2\xi}{\delta z^2}+c\frac{\delta^3\eta}{\delta z^3},\quad \frac{\delta^2\eta}{\delta t^2}=b^2\frac{\delta^2\eta}{\delta z^2}-c\frac{\delta^3\xi}{\delta z^3}.$$

This addition was justified by his obtaining equations for the motion of the two rays representing right- and left-handed circular vibrations, and his resulting expression for the rotation,

$$R=\frac{2\pi^2cV^2}{b^4\lambda^2},$$

shows it to be inversely proportional to square of the wave-length λ, if V, the velocity of the ray within the medium, is constant, that is to say, if dispersion is neglected. The constant c, which is shown to be very small, is positive in a dextro-rotatory, and negative in a laevo-rotatory crystal. The full investigation is to be found in Verdet's *Leçons d'optique physique*, vol. ii., together with a bibliography of other writings on the subject published prior to 1866. A more modern treatment by Drude, based on the electromagnetic theory, is contained in the latter's text-book on Optics which has already been mentioned.

A. B. D.

Polarised Light, Methods of Production of. See "Polarised Light and its Applications," § (8).

Polish of an Optical Surface, discussed by Lord Rayleigh. See "Optical Parts, The Working of," § (5).

Regarded by Beilby as a rearrangement or glow of the surface molecules. See *ibid.* § (5).

[1] *Camb. Phil. Soc. Trans.*, 1831, iv. 79.

Polishing, Times of, with typical polishing materials under specified conditions, tabulated. See "Optical Parts, The Working of," § (8).

With a variety of materials, when using a particular polishing medium, under specified conditions, tabulated. See *ibid.* § (8).

Polishing Materials. See "Optical Parts, The Working of," § (8).

Polishing Optical Surfaces, Tools for. See "Optical Parts, The Working of," § (9).

Portable Photometers. See "Photometry and Illumination," § (61) *et sqq.*

POSITION-FINDER, THE MIRROR

§ (1) General Principles.—In the development of the theory and practice of anti-aircraft gunnery it became necessary to employ some rapid and accurate means of determining the path, or the position at any moment, or the velocity, of an object in the air. To meet this need the mirror position-finder was designed early in 1916. This instrument, however, gave satisfactory results only at high angles of elevation, above (say) 35°, and at the end of that year the window position-finder was designed to deal with the lower angles. The principle of either instrument is the same, and together they make it possible to record the positions or movements of aerial objects, from the vertical to the horizontal. The principle is illustrated most simply by *Fig.* 1, in which GK, G'K' represent two rectangular sheets of glass lying in the same plane, but at a considerable distance apart. At A_1, A_1', are two small apertures through which observers can look at a distant object, P, on the further side of the glass. Each aperture is at the same distance, h, from the glass, and the distance, B, between A_1 and A_1' is known accurately. Perpendiculars from A_1 and A_1' meet the glass at O and O'. The observer at A_1 marks on the glass at m the apparent position of the object, P; the observer at A_1' does the same at m'. If the positions of m and m' relative to O and O' are determined, the two lines A_1m and $A_1'm'$ are fixed and a simple calculation gives the position of P.

If the glass sheets are horizontal and the eye apertures are placed as shown at A_1 and A_1', it is necessary for the observers to look upwards. This is inconvenient, and the difficulty is avoided in the mirror position-finder by the use of horizontal mirrors. The apertures, A and A', are placed above the mirrors, each at a height h, and in such a way that A_1 is the image of A and A_1' that of A'. The observers look downwards at the images of P in the mirrors. In the window position-finder the glass sheets are vertical

and the observers look through them in the manner described above.

§ (2) Description.—The mirror position-finder consists of two horizontal plane mirrors, etched on the silvered side in centimetre squares, mounted—if possible at the same level—at the ends of a base of measured length B, aligned accurately on one another and carefully levelled. The stations (which we will call O and O′) are connected by telephone; the observer at each station wears a receiver on his head and holds a microphone in his left hand. Each mirror is provided with a movable stand, carrying an aperture

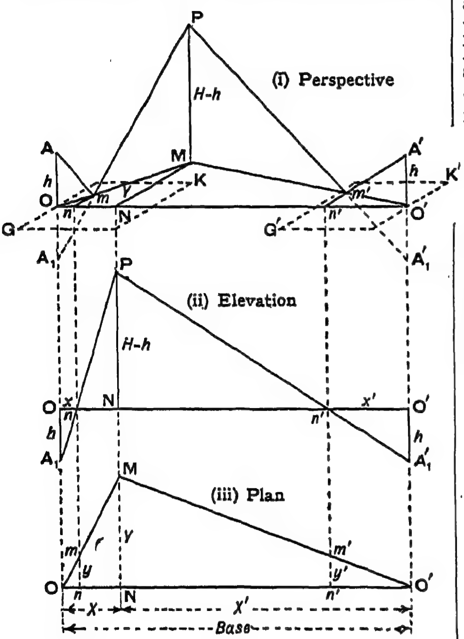


Fig. 1.

at height h above the mirror: it is best to employ a skeleton aperture so as to impede the view as little as possible. The observer looks through the aperture at the reflection in the mirror of the object which he desires to observe, and notes with a pen on the surface of the glass the position of the image at any desired moment or series of moments, care being taken to put the real point of the pen, and not its image, on the image of the object. Simultaneous observations are made at both stations, and from these observations the position of the object can be simply calculated. The ink used should be of the non-drying variety (made with glycerine), both to ensure that the pen marks when required and to enable a permanent record to be made of the observations. For the latter purposes, after the observations are completed, a pair of suitable axes is inked in on the glass, and a piece of absorbent ("Roneo") paper laid over it. The results can then be measured up and the appropriate calculations made subsequently.

The aperture is placed vertically above the intersection of two suitable rulings on the glass, by looking through it and placing its image on that intersection. The positions of the dots recording the position (or rather the direction) of the object are expressed in Cartesian co-ordinates, measuring from the point on the mirror vertically below the aperture, x being the distance in centimetres parallel to the base, and y the distance perpendicular thereto, x' and y' being corresponding co-ordinates of the dot made at the other station O′. The position of the object in space is expressed similarly in Cartesian co-ordinates and referred to the same two axes, being $\{X, Y, H\}$ as measured from O, and $\{X', Y', H\}$ as measured from O′. It is assumed that X, X', x, x', are all measured *towards* the other station, that the co-ordinates of the object in space are very large compared with those of its image in the mirror, and that the stations are at the same level. It is obvious then from considerations of similarity (see *Fig.* 1) that the following relations hold:

$$\frac{X}{x}=\frac{Y}{y}=\frac{H}{h}=\frac{X'}{x'}=\frac{Y'}{y'}=\frac{X+X'}{x+x'}=\frac{B}{x+x'},$$

since $\qquad X+X'=B.$

Hence $\qquad X=\dfrac{xB}{x+x'},$

$$Y=\frac{yB}{x+x'},$$

$$H=\frac{hB}{x+x'},$$

and $\qquad y=y'$, since clearly $Y=Y'$.

The formulae for X, Y, and H enable the position of the object in space to be calculated by simple arithmetic and without the use of trigonometry. The relation $y=y'$, on the other hand, is a valuable check on the observations, and may be employed also to determine simultaneous points on two continuous records (one at each station) of the path of an object in space. Indeed the "equality of the y's" enables the complete track of an aeroplane or other aerial object to be calculated without any means of ensuring simultaneity of observation, the two complete records being co-ordinated with one another merely by taking points with equal y's. The method, of course, breaks down when the path of the object is parallel, or nearly parallel, to the base. The "equality of the y's" has a third value in

enabling the observer at one end of the base to point out objects to the distant observer, so enabling him, for example, to discriminate between different aeroplanes, or different shell-bursts, or different parts of a cloud.

§ (3) THE OBSERVATIONS.—The method of observation is not difficult: it involves one of the most refined and highly developed of human faculties—the accurate adjustment of hand and eye—and so enables the record to be made with quickness and precision. The rulings on the glass are a great assistance in accurate observation, by enabling the eye and brain to "fix" the direction of the object at any desired moment, and so to note it with the hand without undue hurry. With every precaution taken, good glass for the mirrors, telescopes for alignment, sensitive spirit-levels for levelling, fair visibility, and trained observers, experience at Portsmouth during the war in recording the positions of shell-bursts in the air showed that the probable error of one observation was not more than 0·1 or 0·2 mm. in x or y, an error corresponding usually only to a few feet in the calculated position of the object. The most serious error indeed, unless special precautions be taken, is caused by unevenness in the glass, which should be chosen (or better, manufactured) as flat and as parallel-sided as possible.

§ (4) GEOMETRICAL PROPERTIES. — The mirror position-finder possesses several useful geometrical properties in addition to the "equality of the y's." For example, if, as is common, an aeroplane be flying at a constant height the track on either mirror is exactly similar to the track in space, reduced in the ratio h/H. Or again, the height is given by the simple formula $\mathrm{B}h/(x+x')$, thus it is inversely proportional to the sum of the x's. Or again, the velocity of a horizontally moving object (such as a puff of smoke in the wind) is exactly the same as that of its image, recorded at measured intervals on the glass, multiplied up in the ratio of H/h. If, therefore, H be known, as, for example, in the case of a shell fired to burst at a known height, the speed and direction of the wind at that height can be determined at once merely by recording the motion of the image in one mirror of the smoke from the shell.

§ (5) THE WINDOW POSITION-FINDER. — This is similar to the mirror instrument in principle, the whole system of glass and aperture being revolved through 90° about the base line, the mirror in this case being replaced by ruled transparent glass, and the observer noting the direction of the object in ink on a vertical window instead of on a horizontal mirror. The practical working details of the instrument are different, and observation is not quite so easy or accurate, but the theory and the calculations remain the same. The geometrical peculiarities of the record of an object flying at a constant height of course no longer obtain. The instrument may be used for the accurate observation of objects or events in the sea or on the ground, as well as for aerial observation.

§ (6) USES OF THE POSITION-FINDER. — The instruments described have been employed for a variety of purposes:

(1) For determining the height, position, path, or speed of an aeroplane flying horizontally or in any manner required.

(2) For determining the position in space of a shell bursting in the air.

(3) For determining the point of impact of a shell striking the sea.

(4) For determining the velocity and direction of the wind at any height desired, by firing a shell to burst at that height and making timed observations on the path of the smoke; in this way observations of the wind have been made at heights up to 35,000 ft., and of speeds up to 115 miles per hour.

(5) For determining the height, direction of motion, and speed of a cloud.

The mirror and window position-finders have been and are employed by the British military, naval, and flying services for gun trials of various kinds, for recording anti-aircraft practices, for determining the height, speed, or position of aircraft, and for measuring the velocity and direction of the wind at various heights. The U.S. Army also have erected a number of stations at their Aberdeen Proving Ground in the Chesapeake Bay.

A. V. H.

POSITION FIXING, METHODS OF, for ships at sea. See "Navigation and Navigational Instruments," § (17).

POSITIVE RAYS: a stream of positively charged atoms travelling mainly towards the cathode when an electric discharge is passed in an evacuated tube. See "Radiology," § (5).

POT FURNACES FOR MELTING GLASS. See "Glass," § (15) (i.).

POTASH, USE OF, IN GLASS MANUFACTURE. See "Glass," § (5) (ii.).

POTASH LEAD GLASS, presence of chlorides or sulphates leads to cloudiness. See "Glass, Chemical Decomposition of," § (1).

POTASH LIME GLASS. See "Glass, Chemical Decomposition of," § (1).

POWER OF A LENS, the reciprocal of its focal length. For methods of determination see "Lenses, The Testing of Simple."

PROJECTION APPARATUS

§ (1) GENERAL.—Projection apparatus may be broadly defined as apparatus in which a source of light is associated with an optical device to produce localised distant illumination. Most light sources send out light more or less equally in all directions, and it is evident that if an optical device can be employed to bend the rays all into one direction, it should be possible to increase the volume of light passing in that direction enormously, and so ensure a degree of illumination at a great distance in that direction only, which would be equivalent to the illumination produced by the naked source alone at points comparatively near to it.

If the theory of projection apparatus is to be understood, it is necessary to investigate, first, the general theory of illumination, and secondly the theory of optical devices for bending the ray paths. Now the broad principles underlying both these are quite simple, and yet there is the profoundest ignorance as to the limitations of projection apparatus amongst otherwise well-informed people; such ignorance, for instance, as caused numberless inventors in all seriousness to propose that airships could be set on fire by concentrating the rays of a searchlight upon them. The object of this article is to explain the broad principles underlying the theory and practice of projection apparatus in the simplest possible way.

To begin with the theory of illumination; this obviously involves the measurement of brightness, and though accurate measurements in the laboratory are made with the aid of an instrument called a photometer, yet this is only a mechanical device to assist the eye. Ultimately the eye is the instrument that is used to compare one brightness with another, and the essential theory of illumination can best be discussed from the point of view of *what the eye sees* at any given point.

To be exact, no human eye can verify the illumination produced by a source of light *at any given point* of space; it can only form an estimate of the total amount of light energy from that source passing through the iris aperture, which has very definite though variable size.

For results to be strictly comparable, therefore, a "hypothetical eye" must be assumed with a small fixed iris-opening, capable of examining and estimating the illumination produced by the brightest sources without fatigue or dazzle. It will also be convenient to assume that this "hypothetical eye" has an infinite capacity for seeing detail so that it would recognise the shape of even the most distant sources. In what follows the term "eye" will be used as signifying an organ with these extended powers.

§ (2) DEFINITION.—It is desirable at the outset to get rid of that meaningless abstraction a "point of light." There is no such thing as "a point of light." Light cannot be obtained except from an incandescent source of definite and measurable size, and the theory of illumination can never be understood until this elementary fact has been clearly grasped. Light is a form of energy which is emitted from incandescent surfaces. Any particular surface will only emit visible light when raised to a certain minimum temperature; and after that temperature has been passed, it will emit more and more light as the temperature is raised. Surfaces made of different materials, but otherwise identical, when raised to the same temperature will usually emit quite different amounts of light per second.

The term "intrinsic brightness" is usually applied to the measure of this light-emissive power per unit of area of surface, and it is generally quoted in candle-power per square inch.

The fundamental fact on which the whole theory of illumination depends is that each element of any surface raised to a uniform temperature, and emitting light in consequence, appears always of the same brightness at whatever angle or from whatever distance it is viewed. Thus a uniformly bright surface appears to the eye simply like a flat sheet of brightness having a certain definite "apparent size and shape." For the sake of clearness

the "apparent size and shape" for any "view point" may be defined as the actual size in square inches and shape of a flat sheet which, when placed one foot from the eye with its plane perpendicular to the line of sight, exactly obscures the incandescent source or appears to coincide with it in shape and size as seen from that "view point."

It follows that the illumination produced at any point is proportional to the "intrinsic brightness" of the incandescent surface and to its "apparent size," since 2 square inches of flat glowing surface at 12 inches distance must produce double the illumination of 1 square inch of surface glowing with equal brightness at the same distance.

The unit of illumination is known as a foot-candle and is the illumination produced by an incandescent sheet of unit "intrinsic brightness" at a point at which it has unit "apparent size."

Again, the same sheet raised to a higher temperature will appear brighter; in fact, the "intrinsic brightness" of any particular material is a measure of its temperature; but different substances raised to the same temperature will glow with different brightnesses. The illumination produced, however, at any given standard distance can be made the same by increasing the *size* of the less bright material so as to compensate for its smaller "intrinsic brightness."

While, however, the illumination produced by a large gently glowing sphere may be of exactly the same intensity as that produced by a tiny intensely bright sphere, yet the two sources act *very differently* when associated with an optical device for projecting a beam localised in a certain direction.

§ (3) Optics of Illumination.—It is the peculiar function of optical devices that they can entirely alter the apparent shapes and sizes of sources as seen from certain directions, but they can never make them look brighter. Owing to transmission losses—colour in glass, or imperfect reflective power—they may make the source look less bright, but the variation in intensity of illumination produced in certain directions by an optical projector is simply due to its capacity to make the source look larger or smaller in those directions. Thus, an incandescent sphere S (*Fig.* 1) placed behind a plano-convex condenser may appear when seen through the lens L as of apparent size E instead of *e*, which would be the apparent size of the source as seen through a thick plane plate of glass in place of the lens L. The brightness of the source as seen through the parallel plate is exactly the same as the brightness of the magnified source when seen through the lens instead of the parallel plate, and this brightness is necessarily slightly less than the brightness of the naked source, owing to absorption and transmission losses in passing through the glass.

From the foregoing it is evident that there is a very definite limit to the intensity of illumination that can be produced by any projector of given size using a source of given "intrinsic brightness." The very highest efficiency for such a projector is that the whole front window or aperture of the projector, as seen from a distant point, shall be filled (or flashed) with the "intrinsic brightness" of the source. This is termed a "complete flash."

When only certain parts of the front aperture are seen as of the same "intrinsic brightness" as the source (less ordinary transmission losses), then the projector is said to afford a "partial flash." This latter term, it should be noted, applies not only to systems leaving dark patches in the front aperture, but also to systems in which certain areas of the front aperture are seen filled with a *coloured flash*, due to the "partial flashing" of certain primary colours causing such constituents to be absent in those areas.

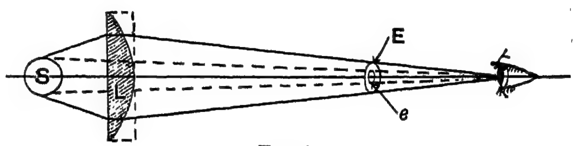


Fig. 1.

It will be noted that the condition for a "complete flash" will be satisfied if the ray directions traced from the eye through every point of the optical system are so bent and deflected that they all concentrate on to the source. Since a *distant* view point is assumed, such ray directions form a system of practically parallel lines at the projector. It will be seen, therefore, that for maximum efficiency with a source of minimum size the optical system must be designed to bend parallel ray directions, so that they all pass through, or at least *extremely* near to, a fixed point termed the "focus."

If then a small source be placed at that "focus" there will be one direction in which a relatively powerful beam will be projected. If, however, a source of the same candle-power but $\frac{1}{100}$th the apparent size, as seen, from the projector, be placed at the "focus" of the same projector, since its "intrinsic brightness" must be 100 times as great to make it the same candle-power, the beam projected will be 100 times as powerful, always provided the projector is efficient

enough to afford a "complete flash" in both cases. If, however, the second source be increased in size while remaining of the same "intrinsic brightness," no increase in the illumination produced in that particular direction will take place as a result of such increase in size, however great it may be, because the same-sized optical system affords a "complete flash" in either case.

It will naturally be asked, "Where then does all the increased light energy go to, if the intensity of the beam remains unaltered when a source of much greater candle-power is substituted?" The answer is that the beam projected fills a greater angle, so that if thrown on a distant white screen it will afford a "light patch" of bigger area. For all view points on the screen at which the optical system affords a "complete flash," the illumination produced will be the same, whether the source be large or small, always assuming the same "intrinsic brightness" for the source.

It is convenient to investigate projected beams by noting the shape and general intensity of the "light patch" projected on a distant white screen, and so it will make the argument clearer if the projected beams are considered in terms of the "light patch" received on a perfectly efficient white screen placed at some definite standard distance from the projector—a distance that shall be very large in comparison with the size of the projector. At whatever angle it is viewed, such a screen would make the brightness of the "light patch" at any point always strictly proportional to the illumination produced at that point by the projector. Each element of such a screen covered by the "light patch" would act like a self-luminous surface and send out light equally in all directions.

If, instead of the flat screen placed at standard distance from the source, a large hollow spherical screen of some definite standard size be imagined extending all round the source with the source at its centre, it will be seen that the optical projector, whatever it is, can only make use of the light energy represented by the complete spherical "light patch" cast by the original source on this spherical screen. Whatever light the optical system can intercept and bend in the required direction will be removed from the light falling on one part of the sphere and added where it is wanted. The most, therefore, that any optical system can do is to bend the rays of light in such a way that all the light energy spread over the whole sphere by the naked source is concentrated so as to fall only on one small portion of it.

Evidently, therefore, while optical projectors may be designed to produce relatively intense illumination from sources of small candle-power, yet the angular size of the beam projected under such circumstances can only be very small.

If the source were a glowing sphere the illumination produced by it in every direction would be the same, and if it were desired to produce from such a source by optical means a circular "light patch" one-thousandth of the area of the complete spherical surrounding screen (this would correspond to a beam having a semi-angle of about $3\frac{1}{2}°$ or a total angle of divergence of about 7°), then it would naturally follow that the most that could be expected would be that the intensity of the projected beam would everywhere be one thousand times as great as the intensity due to the naked source alone.

It would follow, therefore, from what has gone before that the front aperture of the projector must have at least one thousand times the "apparent size" of the spherical source as seen from the screen, since it is only by increasing the "apparent size" of the source by optical means to fill this aperture that the increase in illumination can be obtained.

It is always impossible to collect from all round the source, and it is generally only convenient to collect from a comparatively small fraction of the total area of the surrounding sphere. Twenty-five per cent of the complete sphere, or 50 per cent of the hemisphere, for instance, is quite good for a searchlight. If this percentage only be collected and it is desired to include just the same angle in the projected beam as before (viz. 7°), then it would only be possible to get one-fourth the light in the beam using the same imaginary spherical source and designed as before for "complete flashing." This projector, therefore, would have to have half the previous diameter; or if the *same* illumination were desired filling the same angle, then the source assumed to be of the same "intrinsic brightness" would have to have four times the area, *i.e.* have twice the diameter.

Where the optical system is symmetrical about an axis, the "angle of collection" is quoted as the angle of the right circular cone formed by the extreme rays collected by the projector.

To sum up, therefore, from the foregoing general considerations, it will be seen that if the *type* of source is given (*e.g.* acetylene flame, oxyhydrogen limelight, electric arc), in other words, if the "intrinsic brightness" is given, then the required intensity in the beam can be secured by having the projector big enough. On the other hand, a wide angle of divergence in the beam will be secured either by collecting the light from as large an angle as possible

or by employing a source of as large a size as possible. Neither of these two latter conditions, however, affects intensity, which is solely a question of making the front aperture as large as possible; it being always assumed that the optical system is equally efficient in every case as regards its ability to afford a "complete flash."

Again, if by any means the "intrinsic brightness" can be increased fourfold, then exactly the same beam having the same intensity and including the same angle can be projected by using a projector one-half the diameter in conjunction with a similarly shaped source only half as large (its total candle-power being the same, since it is assumed to have four times the "intrinsic brightness"): by increasing the "intrinsic brightness" sixteen times, the whole apparatus could be made one-quarter the diameter and yet give the same intensity within a beam of the same angle, and using a source of the same candle-power.

As these general considerations are very important, they will be again summed up in the following manner. The factors involved in the projection apparatus are:

A. The "intrinsic brightness" of the source.

B. The size of the source measured by *area*, or its apparent size, at standard distance.

C. The fraction of total energy of source collected by the projector.

D. The size of the projector, which must be measured by the *area* of its front aperture.

The resultant "light patch" projected has:

I. Size measured by its area.

II. Intensity measured by its brightness.

If one factor alone be varied, while the others are kept constant, it may be said:

Varying A causes II. only to vary proportionately.

Varying B causes I. only to vary proportionately.

Varying C causes I. only to vary proportionately.

Varying D causes II. to increase and I. to decrease proportionately to increase in D.

It should be noted that in the above the same efficiency for the projector is always assumed, that is to say, it need not afford "complete flashing" in every case, provided it always affords the same percentage of its front aperture flashed. It will be shown later that it is very difficult to increase the angle of collection of a projector without lowering its efficiency, so that when C is varied a fresh factor is introduced which modifies the above result.

§ (4) Projection Apparatus.—The study of projection apparatus is now seen to involve the study of the flashing of incandescent sources as viewed through optical devices.

The directions in which the eye sees things through an optical system can be ascertained mathematically by what is termed "ray tracing," since at every point of reflection or refraction in the optical system the change of direction follows known laws. If, therefore, a distant view point be taken and all possible ray directions emanating from it be traced through the optical system, then for well-designed projection apparatus they should, after being so traced through the optical system, come, if not to a true point focus, yet to a very high degree of concentration near the "focal point."

If the ray directions from a distant point on the axis of an optical system, traced through that optical system, all converge to a "point focus," then any source, however small, which includes that "point focus" within it, will be seen from that distant point as completely flashed. That is to say, an extremely diminutive source placed at the "focal point" would send a very powerful beam along the axis, though it would naturally have an extremely small angle of divergence. This "point focus," however, is impossible of attainment, though the ray directions do, as a matter of fact, come to a remarkably high degree of concentration at the focus of well-designed and well-made apparatus. If the smallest possible sphere be described surrounding what may be termed the "focal point" so as just, and only just, to include all the convergent rays, either intersecting it or just touching it, then this may be termed the "focal sphere."

When the source actually employed in projection apparatus is not larger than the "focal sphere," "partial flashing" alone is possible, and the theory is rather more difficult; also it is extremely difficult to get uniform brightness in the "light patch."

In the great majority of cases it is desired to project a beam which shall be symmetrical about an axis, and the optical system employed is also symmetrical about an axis. In fact, the difficulty of making optical devices which are not symmetrical about an axis is such that when unsymmetrical beams are wanted they are usually obtained by combining a symmetrical optical system with a suitably shaped source.

Optical systems symmetrical about an axis (in future termed "symmetrical projectors" for shortness), however complicated they may be and however often the rays passing through them may be bent or reflected, have the following very interesting properties which are of great value in elucidating the theory of their use for projection purposes.

If a ray PR (*Fig.* 2) drawn parallel to the axis SF of a symmetrical projector be traced through the optical system, it must always

on emergence intersect the axis at some point F. Again, if the finally emergent ray be produced backwards to meet the line of the

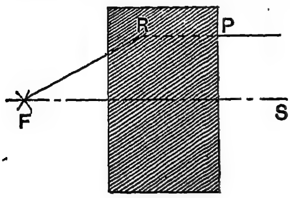


FIG. 2.

incident ray in R, then the "apparent width" of any small source (measured perpendicular to the axial plane PRF) as seen through the projector from P will always be the same as the "apparent width" of the naked source as seen directly from R (taking width again to refer to directions perpendicular to the axial plane PRF).

The "apparent height" of the source as seen from P may be either magnified or minified by the optical system as compared with the "apparent height" as seen from R; but all that can be said for certain is that the angular value of the width of any small source as seen from P through the optical system is exactly the same as the corresponding angular width of the naked source as seen from R; the whole "apparent shape" is similar, but possibly distorted, either drawn out or compressed in the direction of the axial plane containing the "view point."

If P be a point on the front aperture of the projector, then the "light patch" which would be projected if the whole front aperture were masked off except a tiny element at P would obviously have the same "apparent shape and size" as the "apparent shape and size" of the source as seen from P through the projector. But this is similar to, and of exactly the same angular width as, the "apparent shape and size" of the naked source as seen from R.

If the projector is to afford a "complete flash" throughout the whole of the projected beam, every such element of the front aperture must project a "light patch" which shall be coincident with that afforded by the complete projector.

Hence, there are two conditions to be satisfied for "complete flashing" throughout the entire projected beam: first, that the parallel ray directions such as PR must for all zones of the projector converge to one definite "point focus" F; secondly, the "apparent width" of the source as seen from every such point as R must be constant, *i.e.* the length RF must be constant for all zones of the projector.

The point R is conveniently termed the "equivalent bending point" for the ray PR, and all such points will lie on a surface coaxial with the optical system. This surface may be termed the "equivalent bending surface."

From the above it is evident that the two conditions to be satisfied by any projector designed to afford "complete flashing" throughout the whole of the projected beam are: first, that ray directions emanating from a "point focus" shall all be rendered accurately parallel to the axis; secondly, that the "equivalent bending surface" for such rays shall be a sphere having the "point focus" as centre.

§ (5) DETAILED CONSIDERATION OF USEFUL SOURCES.—It has already been made clear that the only way to increase the intensity of the beam projected from really efficient apparatus of given size is to increase the "intrinsic brightness" of the source; and hence, sources of the highest possible "intrinsic brightness" must be chosen for powerful projectors.

The subjoined table gives the "intrinsic brightness" of some well-known sources in candle-power per square inch:

Paraffin flame (enclosed in glass chimney)	10
Acetylene flame (burning in air) .	36
Incandescent oil (petroleum) with mantle	340
Ordinary tungsten filament (*in vacuo*)	960
Oxy-acetylene with pastille (in Messrs. Chance's projectors)	4,500 to 5,700
Tungsten filament in argon (·65 watt per candle)	7,500
Tungsten arc or "Pointolite" .	12,000 to 16,000
Tungsten filament in argon (·4 watt per candle)	17,000
The crater of ordinary carbon arc .	110,000
The sun at noon, summer, in British Isles	600,000
The sun at zenith	1,000,000

These figures indicate how hopeless it is to expect to do anything in the way of projecting beams which the eye can see at considerable distances in bright daylight except by using an arc lamp or a very highly overrun (so-called) half-watt electric lamp.

The obvious difficulty in the case of the latter source is the awkwardness of the shape of the filament. This can, however, be wound as an exceedingly close spiral coil, and so be made equivalent to a continuous incandescent cylinder.

The illustration (*Fig.* 3) is from photographs showing an actual incandescent filament eight times natural size, wound so as to form two such incandescent cylindrical shapes with axes parallel and very close together.

Both the "Pointolite" and the tungsten filament lamp are subject to the great disadvantage that the glass containing bulb has

to be very large in comparison with the dimensions of the source itself. This places a very definite limit to the possible proximity

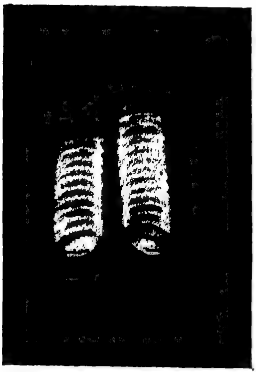


FIG. 3.

of any optical device used for projecting a beam from the source. Consequently such sources can only project beams of small divergence, unless "partial flashing" is resorted to and the projector made unnecessarily large. Another disadvantage is that when used with a reflector placed behind the source the glass containing bulb practically acts like an opaque body to the reflected beam, and so the only effective part of the reflector consists of the outer annulus, the beam from which just clears the bulb.

§ (6) THE PARABOLIC REFLECTOR.—The well-known property of the parabola, viz. that rays A*a*, B*b*, C*c*, D*d*, etc., drawn parallel to the axis FX (*Fig.* 4), are reflected at the curve so

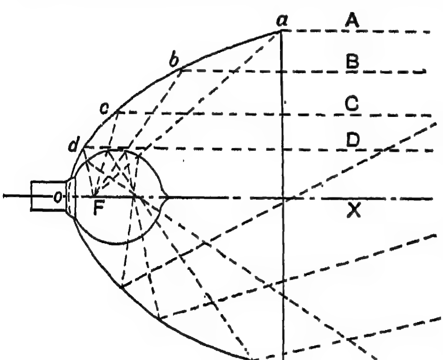


FIG. 4.

as to pass through the focus F, makes the parabolic reflector the most obvious optical element to employ for projecting powerful beams.

Inventors of special projection devices, however, have tended to overlook two important considerations, the first being that a parabolic reflector with a true point focus is a pure mathematical abstraction, and any commercially produced parabolic reflector has a "focal sphere" [1] of very definite and appreciable size.

The second is that the parabolic reflector only satisfies the first condition for affording a "complete flash" with an abruptly defined margin; the second condition, viz. that the "equivalent bending points" [2] shall all be equidistant from the focus, is obviously not satisfied by any parabolic reflector having any appreciable "angle of collection."

Thus, owing to the rapid increase in the focal distances to the points of reflection, as the vertex of the mirror is departed from, any parabolic reflector used in conjunction with a source larger than its "focal sphere" will project a beam in which the marginal zones will be weak.

§ (7) THE PARABOLIC REFLECTOR AND ELECTRIC HEADLIGHTS FOR MOTOR-CARS.—There is a particular type of projector, however, for which the parabolic reflector used in conjunction with an electric incandescent lamp is admirably adapted, viz. the motor-car electric headlight.

What is wanted here is a very wide angle beam, but of very much greater intensity within a central narrow angle as compared with the intensity of the marginal portion of the projected beam, and this kind of distribution is easily obtained from any parabolic reflector having a very large "angle of collection." For motor-car headlights the "angle of collection" is so large that, as shown in *Fig.* 4, the only light not collected from the source passes out from the front aperture of the lamp and forms part of the useful wide angle beam.

The "equivalent bending surface" [3] *is* the surface of the parabolic reflector, and the focal distances *a*F, *b*F, *c*F, *d*F vary enormously; in fact, with the angle *a*F*a′* equal to a right angle, the length *a*F is necessarily nearly seven times as long as the focal distance *o*F, and the whole front aperture *aa′* is necessarily more than 9½ times this same focal distance. But the glass containing bulb of the electric lamp used has got to go inside the reflector, and this sets a very definite limit to the smallness of F*o*, and hence to the possibility of keeping down the size of powerful headlights of this type.

To keep down the size, designers have arranged to place the source low down in the bulb, comparatively close to the cap, thereby increasing the percentage of light lost on the cap and also that lost due to

[1] For definition see § (4). [2] *Ibid.* [3] *Ibid.*

reflections from the bulb, the reflection on going slantwise through the walls of the bulb being much greater than the reflection when going perpendicularly through it. This reflected light, however, is not entirely lost to the beam; some of it goes to augment the wide angle beam, partly directly and partly indirectly, after undergoing a second reflection at the parabolic reflector, as shown in *Fig.* 4.

Starting from *o*, therefore, and going round the source, it will be seen that first there is a fairly large angle over which the light energy is absolutely lost on the cap; then there is a zone in which the light energy which gets through the bulb on to the reflector undergoes reflection at the parabolic mirror, followed by glancing reflection on the electric bulb. This ultimately gets added to the wide angle beam, but it contributes an entirely negligible quantity either to the central intense beam, or to the relatively fairly powerful beam that should closely surround it. It is only when the zones at *c* and *b* are considered that a relatively powerful beam results, since the angle within which the projected beam lies is so small in comparison with the angle around the source from which the "light energy" is collected.

It will, of course, be understood that with a source having an "intrinsic brightness" as high as 960 candles per square inch, anything approaching a "complete flash," even confined to the very centre of the beam, is entirely out of the question. A "complete flash" in a 10-inch headlight would mean a candle-power in the projected beam of more than 37,000, even if the efficiency of reflection be assumed as one-half. This would produce an illumination equal to full moonlight at a distance of well over a mile. No effort is made, therefore, to confine the filament close to one "focal point," nor is the reflector made very accurately so as to afford a very small "focal sphere."

As the bending of the rays in this case is done by pure reflection only, the "light patch" projected by each portion of the reflector is identical in "apparent shape and size" with that of the source as seen from the corresponding part of the reflector. The ray directions drawn in *Fig.* 4 correspond to the centre in each case of the "light patch," which will obviously be the sum total of a lot of thread-like twisted shapes corresponding to all the different views of the filament as seen from *a*, *b*, *c*, etc.

Obviously the zone at *a* is responsible for a narrow angle beam, that at *b* for a wider angle beam, and so on in the inverse proportion of the focal distances *a*F, *b*F, etc.

Commercially obtainable reflectors do not cause rays emanating from a "point focus" F to form a parallel beam after reflection. They have small errors of shape which cause the ray directions to vary within small limits, and provided these limits are kept well within the angular value of the central intense beam (about 6°) no real harm results. These irregularities simply further confuse the superimposed images and cause the reflectors to be less sensitive to displacements of the source from the theoretical focus.

Defective shape is not responsible for the poor performance of many electric headlights seen on the road, but imperfect focussing, or location of the source altogether outside the true "focal sphere." Displacement of the source in an axial direction may result in a hollow beam, there being no light from the reflector in the central part at all.

Displacement of the source perpendicular to the axis, causing an excentric position of the filament, affords an extremely unsymmetrical "light patch" with a very big halo all on one side.

To take a concrete instance: an ordinary candle-power for the bulb of a 10-inch electric headlight is 24. Allowing for light lost on the cap, the parabolic reflector can be designed to collect about 70 per cent of the total light energy emitted, and under fair average conditions it can be assumed to reflect 40 per cent of this. Thus, about 28 per cent is transmitted to the beam, and if one-quarter of this is accounted for by the central 7° wide beam, and the remaining three-quarters by the wide angle beam up to a limit of 40° on either side of the axis, it can be asserted that the candle-power in the central beam will average about 1600, and the mean candle-power in the wide angle beam will be about 32, as produced by the reflector, to which, of course, must be added the 24 due to the naked source, making a total mean candle-power in the wide angle beam of 56. This represents quite good practice. From the figure quoted for the candle-power corresponding to a complete flash in a 10-inch headlight it will be seen that, even in the most powerful part of the beam, this 10-inch headlight will only afford about $5\frac{1}{2}$ per cent of a "complete flash." The dazzle effect of such an intensely bright source as the incandescent filament is responsible for the illusion that the whole aperture of a good electric headlight is filled with a "complete flash" when viewed axially from in front. By looking at it with a good telescope, provided with an optically worked dark glass, it will be found that only twisted thread-like lines of brightness are seen on the surface of the reflector.

The electric headlight, therefore, with parabolic reflector is an example of a projector with a quite large "focal sphere" affording only very "partial flashing," with a comparatively open spiral coil of incandescent filament

confined, when properly adjusted, within the "focal sphere." The parabolic reflector is carried out to include an "angle of collection" of something like 240° to 270°, and the large forward end of the reflector is responsible for the intense narrow angle central beam, the middle zones for the less intense beam immediately surrounding it, and the back part of the reflector behind the bulb is responsible only for scattered light, forming in association with the naked source and reflections from the glass containing bulb the marginal wide angle beam.

§ (8) The Parabolic Reflector and Electric Searchlights.—Searchlights form perhaps the most typical instance of projection apparatus, since it is the function of a searchlight projector to produce the most powerful distant illumination in a predetermined and controllable direction.

To do this a source of the highest possible intrinsic brightness has to be employed, used in conjunction with a highly efficient parabolic reflector, made as large in aperture as possible, and of as high an efficiency as possible.

This means the combination of an arc light with a *glass* parabolic reflector silvered on the back according to the plan shown in *Fig.* 5.

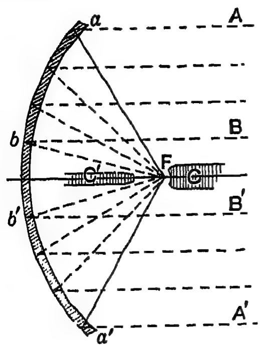


Fig. 5.

The carbons C, C′ are shown disproportionately large, but it will be noticed that the parabolic reflector is carried out to an "angle of collection" of about 120°, and this enables fully 70 per cent of the total light energy emitted by the arc to be collected by the mirror. There is no practical advantage to be gained by trying to exceed this "angle of collection," even if such a mirror could be made to stand the heat of the arc. The negative carbon C′ has its conductivity artificially increased, and is made as small in cross-section as possible, so as to keep its central shadow, *bb′* on the mirror, of minimum size, and for the same reason the arc is made as long as is practicable.

To obtain the great intensity required by modern searchlights, the diameter *aa′* of the front aperture is made in some cases as large as 60 inches, but the difficulty of grinding and polishing, with the necessary accuracy, so large a parabolic shape is so great that there is a tendency for such large projectors to give disappointing results in comparison with the more usual size of 36 inches. The standard focal length for a 36-inch mirror is 17 inches, and this corresponds to an angle of collection of 110°. A useful source to employ with such a mirror would be an arc having a 1¼-inch diameter positive carbon, rated to consume 150 amps. at 60 volts. The brightest part of the crater on such a positive carbon would be limited to a diameter of less than ½ inch. With a source of this size, it would mean that the central zones of a perfect parabolic reflector nearest to it would project a beam of about 1½° and the outer zones a beam of about ¾° angular divergence. The central crater would have an intrinsic brightness of about 100,000 candles per square inch. The mean spherical candle-power within the angle of collection of such a source would be about 18,800 candles, and *if the whole aperture of the mirror were flashed with the intrinsic brightness of the crater* the resultant candle-power would be more than 100 million. But this 100 million is reduced by obstructions such as the negative carbon and the carbon supports and holders. It is also reduced by the losses on going through the mirror, the colour of the glass, which is quite important in such large mirrors, and the imperfect reflective power of even the best glass silvered surface. It is also necessary to protect the arc from the effects of wind; and to do this the front of the projector is glazed with parallel strips of glass. This gives rise to further light losses, and, taken in the aggregate, it can be demonstrated that in actual practice such transmission losses cannot be less than 40 per cent, and may very well be considerably more.

This would give a *maximum possible* intensity of 60 million candles in the projected beam, provided the diameter of the focal sphere afforded by the mirror was really considerably less than the half-inch which is the approximate diameter of the intense central crater on the positive carbon. Such accuracy, however, is not attained in practice, and in consequence the whole unobstructed aperture of the projector is *not* filled with a flash corresponding to the central crater, but parts of it are flashed with a brightness corresponding to the surrounding glowing carbon, which is *very much less* bright. Local errors in shape increase the divergence of the projected beam. In practice they about double the value of the maximum divergence already obtained from purely theoretical

considerations, and further reduce the intensity of the brightest parts of a carefully focussed beam to something like 30 to 40 million candles. Imperfect focussing may still further reduce the intensity of the projected beam, and correspondingly increase the divergence. As a matter of fact an intensity of 60 million candle-power for a 36-inch searchlight has been actually exceeded in practice, by employing specially treated carbons for which an intrinsic brightness of 200,000 candles per square inch has been attained.

It is often stated that the high figures quoted for the candle powers of searchlights are meaningless, because the candle-power is so completely dependent on the state of the atmosphere. Since candle-power represents a *rate* at which light energy is sent out from a source per unit of solid angle, it can have nothing to do with the state of the atmosphere. These fallacious notions are all due to the fact that it is not practicable to compare a searchlight with the 30 or 40 million candles to which (in a predetermined direction and at sufficiently great distances) it is undoubtedly equivalent, whatever the state of the atmosphere.

The state of the atmosphere only comes in when the illumination produced at a very great distance from a searchlight is compared with the illumination produced quite close to a small standard source. It would be very much better if manufacturers of searchlights would give definite guarantees as to what the actual size of the "focal sphere" [1] for their mirrors was, under running conditions, and also their transmission losses. These readily measurable quantities give the best criteria for comparing the efficiency of one projector with another, quite apart from arc efficiency.

Another fallacy to be disposed of in connection with searchlights is the oft-repeated statement that the ideal really aimed at is a perfectly parallel beam projected from a point source. The *idea* is that such a beam would afford constant illumination at all distances, and the divergence of the beam of any practical searchlight is taken as a measure of its inefficiency, because it leads to the falling off in illumination according to the inverse square law.

It is evident that any *efficient* projector must produce a complete flash at all useful long ranges, and hence as the range in any predetermined direction is increased, the falling off in illumination must follow precisely the same laws as would hold for the equivalent candles of that flash, whatever atmospheric conditions prevail.

Constant illumination from a parallel beam at varying ranges requires the use of a source suitably focussed and sufficiently small to afford a partial flash of constant "apparent size." This involves carrying partial flashing to an absurd limit.

[1] For definition see § (4).

In all that has gone before, the range has been assumed to be great in relation to the size of the projector; in fact, the whole theory has been built upon the assumption that the distant view points considered are such that ray directions from them form a system of practically parallel rays over the whole front aperture of the projector.

Searchlight projectors, however, are so large, and the sources used in them so relatively small, that it is quite easy to take fairly distant view points, the ray directions from which form an appreciably divergent system when taken over the whole aperture of the projector. Obviously, the theory breaks down as soon as the departures from parallelism are such that a percentage of the rays, when "traced" through the projector to the source, miss the source altogether instead of ending upon it.

The "focal sphere," however, corresponding to such fairly distant view points is practically identical in size with the true "focal sphere" corresponding to very distant view points; but its position is displaced towards the view point by a very small amount.

Thus, all the characteristics of the "light patch" on a far distant screen can be reproduced on a much nearer screen simply by focussing, *i.e.* by displacing the position of the source by the necessary small amount to correspond with the displaced position of the "focal sphere."

For all ranges outside 1000 yards the focussing displacement for a 36-inch projector is negligible, but to focus on an object at 400 yards' distance requires slightly less than $\frac{1}{32}$-inch focussing displacement of the arc.

Thus, the figure of 30 to 40 million candles found for the beam from a 36-inch searchlight still holds for comparatively near view points if the carbons are properly focussed, but in this case the beam, leaving the lip of the projector as seen by the operator, may very well be convergent.

The beam taken as a whole, however, is not convergent, for the rays cross over after passing the point focussed upon, and form a rapidly divergent beam corresponding to partial flashing at all really distant view points.

As regards the angle of the beam actually projected by practically useful searchlights for long-range work, it usually lies between 2° and 3°, and anything less is not of very much use for searching for objects even at extreme ranges. At three miles range the light patch produced by a 3° beam is about 800 feet in diameter. For comparatively near ranges, however, this angle of divergence gives too

small a width to the light patch projected; consequently, vertical cylindrical elements are fitted to the front window of the projector. The beam traversing these cylindrical elements is spread out evenly in a horizontal direction so that the horizontal angular width of the beam is greater than the vertical angular width by the divergence due to the cylindrical lenses. The cylindrical elements may have a divergence of as much as 30°, and if this were applied to a 3° beam the resulting intensity would naturally be reduced to at least a tenth of its normal value. Instead of affording a complete flash, each cylindrical lens would only show a vertical strip one-tenth of its width flashed; the position of the flashed strip moving across each cylindrical lens from one side to the other as the view point traverses the beam.

§ (9) The Parabolic Reflector and Electric Signalling Lamps.—What is desired in a signalling lamp is to direct an intermittent beam from a person A to a person B so that while B receives a series of flashes as few other people as possible shall receive them.

Obviously everybody in the line AB produced who has an unobstructed view of A *must* see the flashes, but if the beam projected has a very small divergence, few people outside the line AB may see anything of the signalling. At first sight this would appear to call for the perfectly parallel beam affording a "light patch" equal in size to the front aperture of the signalling lamp. It is not so, however, because A who is sending the signal has to *aim* his beam at B by means of some sighting device, and it is impossible to aim a beam with *mathematical* accuracy. When all the possible errors in the lamp are taken into account—errors due to a slightly displaced source—personal errors of the man misjudging his aim—errors due to the uncertainty of the precise location of B and vibration effects due to wind, etc., it will be found that a certain definite though quite small angle of divergence in the beam is required (possibly only 1°). If the signalling lamp instead of being clamped to a stand is held in the hand, and so aimed, it will have to afford a very much wider angle of divergence, say 4°, and if it is to be used as a hand lamp on board ship or on aircraft it must have an angle of divergence of at least 6° to be sure of keeping B within the flash.

For such a lamp, a so-called half-watt electric bulb (tungsten filament incandescent in argon) is a very convenient source. This must be worked at a very high efficiency if any considerable range is to be afforded in *daylight*.

As would be expected, there is an upper limit to the size the projector can have, and if a parabolic reflector is to be used it must only be carried out, as in *Fig.* 6, as far as the *latus rectum*. That is to say, the source (which is at the focus F) must be in the plane AA′ of the front aperture, as the following consideration will show.

It has been shown that the front edge AA′ of the reflector projects the narrowest beam, and with a spherical source the apparent size of the "light patch" projected from this front edge is the same as the "apparent size" of the spherical source as viewed from any

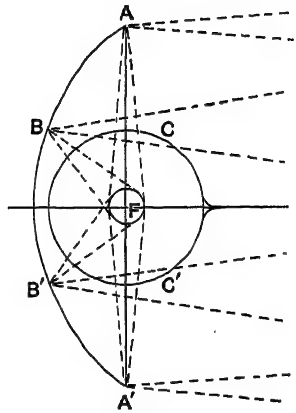


Fig. 6.

point on it. This apparent size is greatest when AF is least, *i.e.* if the diameter AA′ is given, F must be in the plane of the circle AA′.

Of course, the incandescent filament cannot be wound so as to simulate a continuous incandescent sphere, but a very close axial spiral coil (two such coils being shown in *Fig.* 3) will do, if its diameter is made large enough to afford the necessary minimum angle required for the signalling beam, and its length be made long enough, so that the view of the coil as seen from the inner edge BB′ of the annulus actually used for flashing the signal, shall just not show any black centre due to looking right through the axial incandescent cylinder formed by the filament. For reasons already pointed out, the parabolic cap BB′ is practically useless owing to the obstruction presented by the bulb CC′.

For larger and more powerful signalling lamps the electric arc is the best source to employ on account of its very high intrinsic brightness; and electric searchlights, especially in the smaller sizes, are fitted with shutters whereby the beam can be almost instantaneously occulted so as to permit of their being used as daylight signalling lamps.

Since the outer edge AA′ of the reflector is twice the distance from the source that the vertex is, there is a tendency for the beam corresponding to complete flashing to be half the angular width of the whole beam projected.

Such a condition is wasteful since the signalling beam should be equally intense all over, within the limits imposed by aiming considerations, and of zero intensity outside that limit. To secure this, however, it has been shown that it is necessary that the "equivalent bending surface" [1] (which in this case is the parabolic reflector itself) shall be a sphere with F as centre.

§ (10) THE MANGIN MIRROR FOR HEADLIGHTS.—The Mangin mirror is shown in *Fig.* 7, and is a good example of an optical element where the "equivalent bending surface" [1] for parallel axial rays can be made *approximately* spherical with the focus as centre.

Rays $A_1 B_1$, $A_2 B_2$, $A_3 B_3$, etc. (*Fig.* 7) parallel to the axis $A_0 B_0 C_0$ are refracted at the first surface B_1, B_2, B_3, etc., the outer rays being very much more strongly bent outwards than the inner rays, in such a manner

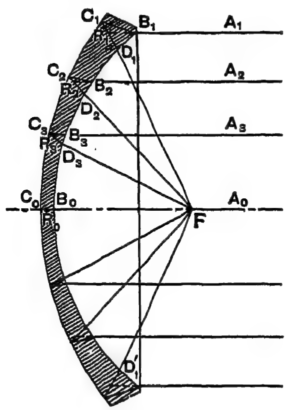


FIG. 7.

that after reflection at the back of the lens mirror at the points C_1, C_2, C_3, etc., they pass almost normally through the first refracting surface again at points D_1, D_2, D_3, etc., *i.e.* they concentrate after reflection to a point F not far removed from the centre of the first surface. An examination of *Fig.* 6 will show that the "equivalent bending surface" determined by the points R_1, R_2, R_3, etc., approximates to a sphere with F as focus.

It is not surprising, therefore, that such a lens mirror affords an extremely abruptly defined "light patch" when used in conjunction with any small axial disc source with its plane perpendicular to the axis of the lens mirror. The "apparent size" of the "light patch" projected is the same as the "apparent size" of the disc source as viewed from the vertex R_0 of the "equivalent bending surface." (This is situated on the axis somewhere within the central thickness B_0, C_0 of the lens mirror.)

[1] For definition see § (4).

For maximum efficiency it is necessary to design the lens mirror so that R_0 is as near to the focus F as the aperture of the mirror permits. As in the case of the parabolic reflector, maximum efficiency is only attained by contriving that the plane of the front edge B_1, B'_1 of the mirror should contain the focus F, but this is not practicable. It is easy, however, to make the angle D_1, F, D_1' as big as 130° or even 140°, and still afford an extremely well-defined "light patch" from any small flat source, and the loss, of course, is not great, because, unlike a spherical source, the flat disc source sends out less and less light at any considerable inclination to the axis; in fact, an infinitely thin flat disc source would send out no light perpendicular to the axis because, when viewed in that direction, it has no "apparent size." The useful "angle of collection" of 140° is too great for such sources as acetylene flames, 120° being about the limit of safe working owing to the tendency of the hot gases rising from the flame to crack the edge of the mirror. The combination of a Mangin mirror, however, with an "angle of collection" of about 120°, and a good flat acetylene flame at its focus, is a very effective piece of projection apparatus for projecting a wide angle beam of uniform intensity. Until the introduction of the electric headlight, this was the combination most in favour for motor cars. A lens mirror with a 10-inch flashing aperture would afford a candle-power in the beam corresponding to "complete flashing" of 1700, assuming 60 per cent reflective efficiency, and with a 36 candle-power flat circular flame, *i.e.* of 1 square inch in area, and an "angle of collection" of 120° for the mirror, this will be spread over a total angle of nearly 13° in the projected beam.

§ (11) THE MANGIN MIRROR FOR SIGNALLING LAMPS.—As has already been shown, signalling lamps require to project a much more powerful beam than motor-car headlights and confined to a much narrower angle. The acetylene flame has to be discarded, therefore, and something like a "complete flash" afforded with a Mangin mirror used in conjunction with a high-efficiency tungsten filament incandescent in argon. The condition to be satisfied is that everywhere within the narrow angular limits imposed for the signalling beam the ray directions traced from the observer's eye through the Mangin mirror to the focus shall terminate on incandescent filament.

For an axial "view point" any annular zone of the Mangin gives a cone of rays which accurately concentrate to a point on the axis. The variation of the position of this point is, in a well-designed Mangin mirror, very small indeed, though not small in relation to the

dimensions of a really closely-wound incandescent coil of reasonably small candle-power. The smallest possible electric source that will give a "complete flash" for an axial view point of any particular Mangin mirror is therefore seen to be a closely-wound axial coil surrounding that short length of the axis within which the "focal points" for all the different zones of the mirror lie.

In such a case the angular limit of the beam is determined by the "apparent size" of the diameter of the coil as seen from the vertex of the "equivalent bending surface" (a point situated somewhere within the central thickness of the mirror).

It should be noted that, as has already been shown, the flash is not complete except in the annulus surrounding a central area equal to the diameter of the electric bulb, and this useful annulus must be such that from no point on its inner edge shall it be possible to look right through the incandescent cylinder formed by the source.

A well-designed Mangin mirror with a small axial coil conforming to the above condition affords the most efficient portable signalling lamp for daylight work, as the following considerations will show.

A suitably designed Mangin mirror of 6 inches clear flashing aperture and with 130° "angle of collection" can be manufactured so that its "focal sphere"[1] is less than ·05 inch. Thus with such a mirror an axial coil ·1 inch long and ·04 inch diameter will afford a "complete flash" in it. The coil when viewed from its side has the "apparent shape" of a rectangle ·1 by ·04 which equals in area ·004 square inch.

Assuming an "intrinsic brightness" of 15,000 this would give a candle-power of 60, which can be quite safely confined within a 2-inch-diameter containing bulb. This means that an annulus having an area of over 25 square inches will be flashed with an "intrinsic brightness" of say 9000 (allowing 60 per cent transmission efficiency for the mirror), and a maximum candle-power in the beam of over 225,000 will be obtained. Thus it is evident that a well-designed small electric signalling lamp can be made greatly to excel the intensity produced by a relatively much larger motor-car electric headlight.

§ (12) The Mangin Mirror for Searchlights.—From the table of "intrinsic brightnesses" already given, it is evident that, even when full allowance is made for the effects of bulb obstructions, an oxy-acetylene projector would have to be made about twice as large as a high-efficiency projector with half-watt focus electric lamp, to afford the same range. In spite of this disadvantage as regards the relatively low "intrinsic brightness," the oxy-

[1] For definition see § (4).

acetylene-cum-pastille is found to be a very useful source for larger units starting at 12 inches in diameter up to say 2 feet in diameter, for the reason that when a total candle-power in the source of about 150 to 200 is exceeded, all the essential conveniences of the electric incandescent filament lamp begin to disappear.

It is the great advantage of the Mangin mirror that the surfaces are spheres and these can be polished true to shape with *extreme accuracy*. It is quite impossible to polish a parabolic shape to the same accuracy. That is why for the smaller units, where the thickness of the glass does not matter, the Mangin mirror is preferred to the glass parabolic reflector.

Taking a medium size of 20 inches in diameter, it is easy to design a Mangin mirror of this diameter with an angle of collection of about 127° whose "focal sphere"[2] (after making full allowance for colour dispersion and errors of manufacture) shall be less than ·2 inch.

This 20-inch mirror used in conjunction with a pastille ·8 inch in diameter would (assuming 60 per cent for its transmission efficiency and 5000 candle-power per square inch for the intrinsic brightness of the pastille) afford a beam whose maximum intensity would exceed 940,000 candle-power, and whose angle of divergence would be nearly 5°.

As has already been pointed out, however, for the largest and most powerful signalling lamps the electric arc must be used, and though the Mangin mirror is a convenient optical element to employ for the smaller projector, for the 24-inch sizes the thin glass silvered parabolic reflector is the most efficient, and from what has gone before it is evident that a beam having a maximum intensity of something like 15 million candle-power should be given by a 24-inch searchlight, even after making allowance for the obstruction due to the shutter employed for signalling.

§ (13) The Condenser Lens.—When a lens is used to bend the directions of ray paths rather than to produce *images*, it is termed a "condenser lens." Such condenser lenses usually consist of very deeply curved plano-convex lenses, either taken singly or in pairs, with their flat surfaces outside and their curved surfaces adjacent. *Fig.* 8 shows the more

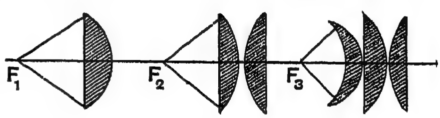


Fig. 8.

usual sections of single, double, and triple condenser lenses with their associated foci F_1, F_2, F_3.

[2] For definition see § (4).

In each case, however, ray directions drawn parallel to the axis are not bent to a true "focal point" or even anything approximating to a point, but the outer rays are more strongly bent than they should be, with the result that in every case the condenser system is found to possess a comparatively large focal sphere.[1]

In fact, a 4-inch diameter single condenser lens having an angle of collection of only 60° would have a focal sphere of ·4 inch in diameter as contrasted with a Mangin mirror which could be made of the same diameter with double the angle of collection and having a focal sphere of only ·03 inch.

For this reason a condenser lens is a bad and inefficient optical device for projecting a powerful beam of moderate divergence from very small bright sources.

Railway signals, however, afford a good example of the successful employment of simple condenser lenses giving wide angle beams of moderate intensity from relatively large and low-intensity flame sources.

The condenser lenses fitted to pocket flash-lamps are not in any sense focussed, and as their flashing is so imperfect, they hardly fall within the category of projection apparatus.

One of the most important uses of condenser lenses, however, lies in connection with the projection of images of transparent objects on a screen.

§ (14) Lantern Projection of Images on a Screen.—The less usual form of lantern projection, viz. the projection of solid opaque objects in their natural colours, will be considered first, because it is the simplest in theory (though the most difficult in practice).

For the sake of simplicity it will be assumed that a sheet of perfectly white paper is laid horizontally at AB (*Fig.* 9) on which perfectly black letters are printed. The lens C placed at a suitable focal distance vertically above it will, in conjunction with a 45° mirror D, project an image A′B′ of AB on to a vertical screen as indicated.

The 45° mirror is needed, because otherwise the image projected would be laterally inverted and the letters would be unreadable except from behind the screen.

The image will only be visible if AB be sufficiently strongly lit and the room be darkened so as to shut off all extraneous light from the screen A′B′ except that coming from the lens C.

To illuminate AB sufficiently strongly, two sources S and S′ may be imagined placed on either side of it and as close to it as possible, giving an average illumination on AB of 20,000 foot candles.

A good "angle of collection" for the lens C would be 14°, which corresponds to a lens of 12 inches focus with an aperture of 3 inches diameter.

[1] For definition see § (4).

Since a hemisphere of 12 inches radius has an area of 905 square inches it is evident

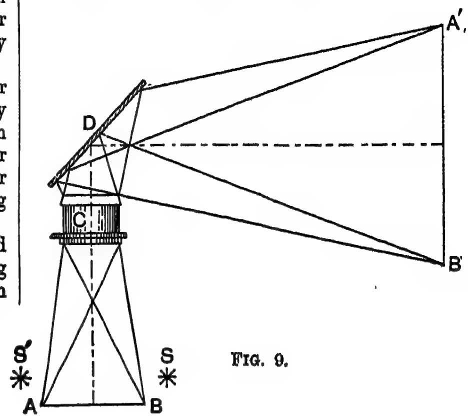


Fig. 9.

that if the perfectly white parts of AB scatter all the light received equally in all directions, the flash in the lens C corresponding to such a white part will have an "intrinsic brightness" of 20,000/905 candles per square inch if no transmission losses are assumed.

Now if the area illuminated on the screen is one hundred times the area of AB, then the apparent diameter of the flash seen in the lens from A′B′ will be one-tenth the "apparent diameter" of the aperture of the lens as seen from AB, *i.e.* it will be ·3 inches diameter at 12 inches' distance. The area of a ·3 diameter circle is about ·0707 square inches, and if a transmission efficiency for the lens C and mirror D of 60 per cent be assumed, the illumination of the screen A′B′ corresponding to *perfectly* white parts of AB is seen to be 12,000/905 × (·0707) foot candles.

This is equal to about ·94 foot candles, and shows how poor the result is, even with an illumination of 20,000 foot candles at AB.

For projecting a transparent object, however, the illuminant can be used in a much more economical manner; for if a source S (*Fig.* 10) in conjunction with a condenser P (usually two plano-convex lenses arranged as in the figure) be so arranged on the axis of the projector lens C so that the image of the source produced by P is concentrated within the lens aperture, it is obvious that any transparent object placed between C and P will, from the point of view of the image of S within the lens aperture, show its clear parts as of the same "intrinsic brightness" as the source S (as seen flashed in the condenser P). Hence, a lantern slide placed upside down at AB, and so that any lettering on it is readable from the same side of it as the condenser P,

will be projected by the lens C so as to be the right way up and with readable lettering provided the focal planes are properly adjusted. But in this case the brightness of perfectly transparent parts of AB is the "intrinsic brightness" of the source S as flashed in the condenser lens, which, of course, is very much greater than that of any illuminated white sheet.

The superiority of the illumination obtained by this method of projection is so great that only when very large magnifications are attempted are lenses of large aperture with the most intense sources needed.

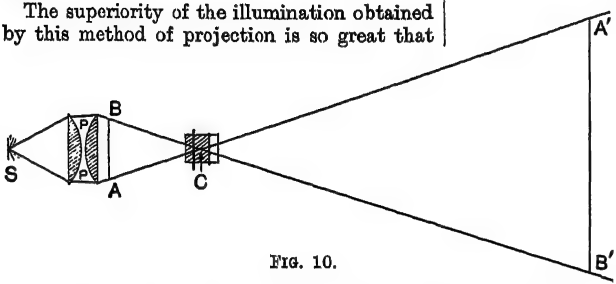


FIG. 10.

Such a case, however, is furnished by the cinematograph projector, where a picture approximately the size of a postage stamp has to be magnified so that its *area* is increased more than forty thousand times. This, of course, calls for the most intense source available, viz. the electric arc.

§ (15) THE CINEMATOGRAPH PROJECTOR.[1]—Since the cinematograph projector must be placed in a fire-proof chamber at the back of the building, its distance from the screen is necessarily great, and in modern picture-houses 90 feet is quite usual. Taking this distance of 90 feet, the width of the picture shown should be at least 17 feet, which is 218 times the width of the film picture and requires the projector lens to have a focal length of about 5 inches.

A rapid projector lens of this focal length might be assumed to have a clear aperture of $1\frac{1}{2}$ inches. Thus, the film picture is actually smaller than the lens used for projecting it, and its area is barely four times that of the crater of the arc used as the illuminant. Lantern slides have been standardised, so that a $4\frac{1}{2}$-inch diameter condenser will easily cover them, and though the problem of the cinematograph projector is so different from that of the ordinary lecture lantern, yet hitherto the same-sized condenser lenses have always been used.

Fig. 11 is drawn correctly to scale one-tenth actual size, and the upper view shows the general nature of the beam projected by a double plano-convex condenser P of 4 inches clear aperture from the crater of an arc S $\frac{1}{2}$ inch in diameter placed 3 inches behind the rear surface of the condenser P, producing a minimum cross-section in the beam at AB (termed the "waist"), of $1\frac{5}{8}$ inches in diameter, 12 inches in front of the condenser. The lower view shows the beam leaving the aperture AB, also the beam leaving the lens C when these are interposed.

The lens C for projecting the image is shown diagrammatically, indicating the clear aperture of the lens elements available for transmitting the beam. The aperture of the extreme rear lens D nearest the gate AB is assumed to be $1\frac{1}{2}$ inches and situated 4 inches in front of AB, the total length of the lens C being assumed as 2 inches.

It is necessary that the "waist" should have a diameter so much in excess of that required to cover the film picture because of the tendency of the arc to shift its position in a variable and uncertain manner; also the illumination of the marginal portions of the "waist" cannot be made as bright as the more central portions.

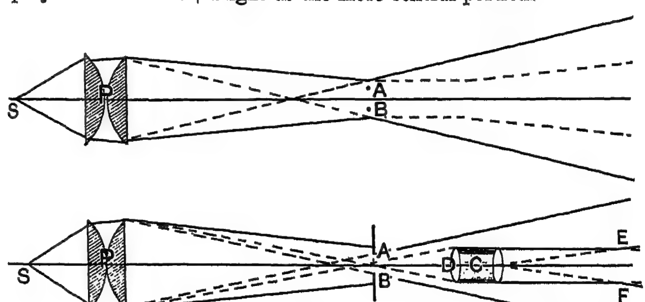


FIG. 11.

It is evident that in no other position in the beam can the film be so intensely illuminated as when situated at the "waist," and this position ensures that all except the marginal parts of the film picture receive a "complete flash" from the condenser.

Consider a small pinhole placed centrally at AB on an otherwise opaque film. The beam from the condenser P passing through this pinhole will cover a $1\frac{1}{2}$-inch diameter circle on the back lens D of the projector lens, and hence will all be collected and transmitted to the screen. If the pinhole

[1] See article "Kinematograph."

be displaced away from the centre, the circular area intercepted by the beam passing through it on to the back lens D moves away from the centre so much that a position is very soon reached in which the projector lens C fails to collect all this light, so that some of it is lost. This loss, however, only affects the marginal portions of the picture.

On the screen the central pinhole will be shown as a white spot, whose diameter will be 218 times that of the pinhole. Since all the light passing through the condenser P to this pinhole also passes through the projector lens C to the screen, it follows that the apparent diameter of the flash seen in the lens C from the screen must be $\frac{1}{218}$ of the apparent diameter of the flash seen in the condenser lens P from the film picture AB.

Thus the illumination of the centre of the screen, in this case corresponding to the whitest parts of the picture, is that produced by a circle $\frac{4}{218}$ inch in diameter (*i.e.* ·000264 square inch in area) at a distance of 1 foot, shining with an intensity of the intrinsic brightness of the arc as reduced by the transmission losses of the whole projector system P and C and the film at the "gate" AB. The mean "intrinsic brightness" of the flash may easily fall below 35,000, allowing for the obstruction of the negative carbon and parts of the condenser not being flashed by the intense central crater of the arc. Multiplying this figure by the above area of ·000264 square inch the illumination of the screen is obtained expressed in foot candles, viz. 9¼.

This result, however, neglects the effect of the shutter employed to cut off the beam whilst changing from one picture to the next. The shutter is periodically interposed where the beam passing from the lens C to the screen is narrowest. This occurs at the image EF of the aperture of the condenser P in the lens C, and in the case under consideration it would be about 6 inches in front of it.

The period of occultation makes the effective illumination on the screen at the most only 60 per cent of what it is with the film stationary and the shutter open. Thus with an ordinary carbon arc 3 inches behind a 4-inch diameter condenser the illumination on the screen corresponding to perfectly transparent parts of a moving picture should be 5½ foot candles. This would give a good bright picture in a well-darkened room.

§ (16) Projection of Microscopic Objects.—This is simply one stage further removed from ordinary lantern projection than the cinematograph projector. Obviously if the linear magnification aimed at is, say, 500 times, the object enlarged can only be very small. But since the apparent diameter of the flash as seen from the screen in the projector lens is only 1/500 of what it is as seen from the object in the condenser lens, this class of work calls for a big angle of convergence from condenser lens to the image of the arc on the object, and a correspondingly big angle of collection for the projector lens employed. The image of the arc produced by the condenser lens need not be large, especially in view of the fact that the field covered by a wide aperture projector lens is small. Consequently the arc need not be placed so close up to the condenser as in cinematograph projection, and a good microscope objective serves very well as a projector lens. The arc not being close up to the condenser makes it possible to use an elaborately designed condenser which will afford a complete flash of the hottest parts of the crater even for the large angle of collection of a good-quality microscope objective. This condition is not satisfied by ordinary condensers, and this explains why the satisfactory projection of highly enlarged views of microscopical subjects involves the careful design of the condenser as well as the projector lens used in conjunction with it.

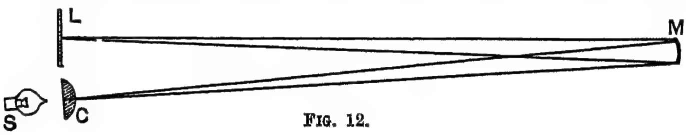


Fig. 12.

§ (17) Scale-reading Projection for Sensitive Instruments.—Under this head must be classed sensitive galvanometers, and all instruments where the angular rotation of the recording element in the instrument is used to make a mirror turn through an angle and so cause the image of an illuminated mark to move through a relatively large distance on a screen with the least possible moment of inertia in the moving parts of the instrument.

The most usual arrangement is shown diagrammatically in *Fig.* 12, and consists of a source S placed behind a condenser lens C, so placed that the image of the source (or more correctly the "waist" of the beam projected from S by C) falls on the concave mirror M, which is necessarily made exceedingly thin and of minimum weight, while a thin opaque vertical line traced on the flat side of C forms an image of itself on the scale L (supposed perpendicular to the plane of the paper). The condenser lens C and the scale L are generally placed equidistant from

the mirror M, whose radius, therefore, is made equal to the scale distance from mirror M to scale L.

Thus a round spot of light is received on the scale screen L with a fine vertical line traced across it. This round spot is the image of the flashed condenser lens produced by M. As usual, however, the question of contrast, or illumination at L, is best settled by placing the "eye" there and noting what it sees. Obviously, for maximum efficiency the whole of the mirror M must be seen as flashed with the "intrinsic brightness" of the source for any point adjacent to the image of the vertical line and as completely not flashed at all for "view points" on this margin. The latter condition can only be satisfied if the mirror M is accurately worked to the spherical shape and the width of the line on C not too fine to allow for the aberrations of the mirror when used at a slight inclination to true normal incidence. The former condition, however, is only satisfied if for all "view points" on the surface of the mirror M the line on C is seen against a background of incandescent surface.

With small electric battery lamps and simple condenser lenses this condition is never even remotely fulfilled, and all that can be provided is that the best and brightest part of the very coloured image of the incandescent filament projected by the condenser lens C is focussed on the mirror M. Thus the eye at L sees a coloured flash in M the "intrinsic brightness" of which is very much less than the incandescent filament.

A device for making sure of using the whole aperture of the mirror is to project on to L the image of a straight portion of the incandescent filament formed by M and to dispense entirely with the condenser lens. In this case, provided only the mirror M is made accurately enough to concentrate ray directions traced from a point on L so that they *all* terminate on such an extremely narrow object (only a few thousandths of an inch) as the filament in a small electric lamp, maximum efficiency will be attained. This accuracy is not, however, possible with commercially obtainable thin galvanometer mirrors, so that in this case, again, partial flashing alone is possible, and maximum efficiency cannot be attained.

The best way to insure maximum efficiency is to make the mirror M (*Fig.* 13) perfectly flat on both sides, because in such forms they can be made much more optically perfect though very thin. The mirror M then simply functions as a device for turning the axis of the projected beam through an angle.

The projector for providing the index mark focussed on the scale L, consists of a separate unit and is arranged on the plan of the cinematograph projector, but using an electric lamp S as a source. The incandescent filament wound as a close spiral coil projects an exceedingly bright image of itself on to the vertical cross-wire N, by means of the condenser lens P, and the image of the cross-wire thus strongly illuminated is projected by the achromatic objective C, so that after reflection at M, it focusses sharply on the scale L.

The sharpness of the image received at L is dependent on the ability of the lens C to afford a complete flash of an exceedingly fine cross-wire at N.

Achromatic objectives, however, are commercially obtainable which will satisfy this flashing test; and provided the mirror M is made accurately flat, this system makes it possible to get a clearly defined index mark on the scale L affording sufficient contrast to be read even in a well-lit room in broad daylight. Moreover, the result is entirely unaffected by the angle turned through by the mirror,

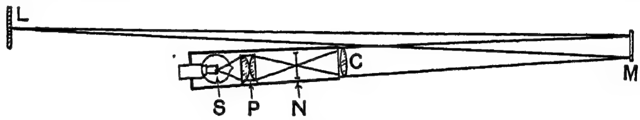


FIG. 13.

whereas a concave mirror has known and traceable aberrations for beams incident upon it at any appreciable angle to the normal or axial direction.

A. C. W. A.

Projection Apparatus, The Study of, involving the study of the flashing of incandescent sources as viewed through optical devices. See "Projection Apparatus," § (4).

Projectors, Photometry of. See "Photometry and Illumination," § (169) *et sqq.*

Pulfrich Refractometer: an instrument for the rapid determination of refractive indices of liquids or solids. See "Spectroscopes and Refractometers," § (13).

Pupil: the circular aperture of the eye. See "Eye," § (21).

Pupilometry: the art of measuring the size and position of the pupil of the eye. See "Ophthalmic Optical Apparatus," § (4).

Purkinje Effect, Influence of, on photometry of surfaces of low brightness. See "Photometry and Illumination," § (124).

Pyramidal Error: the departure from exact parallelism of the edges of a prism. See "Goniometry," § (5).

— Q —

Quantum, Planck's Constant, Values of, obtained by various methods and tabulated. See "Quantum Theory," § (7).

Quantum Limit to a spectrum of X-rays. Planck's quantum relation states that the maximum voltage applied to an X-ray tube multiplied by the electronic charge is equal to the frequency of the shortest wave multiplied by Planck's universal constant. Measuring the limiting wave-length the applied voltage can be found. See "Radiology," § (19).

QUANTUM THEORY

§ (1) **The Quantum.**—It has been found possible to account for many of the phenomena of physics by starting from the dynamical laws of Newton or the electromagnetic laws of Maxwell, together with such extensions of them as relativity theory. Certain phenomena will not, however, fit in, and this in a way that requires not merely a modification of those laws, but some fundamental revolution in our view of nature. The revolution has not yet been made in that no one has suggested any scheme which should cover both the discrepant phenomena and the classical mechanics, but the quantum theory constitutes a first step. At present this theory is hardly more than a working rule, but its wide applicability guarantees that it is leading in the right direction.

The Quantum Theory is definitely contradictory to the laws of mechanics, and this gives an arbitrary quality to some of its applications, in that mutually exclusive ideas are borrowed from both sides and combined. Such a procedure would of course not be tolerated but for the fact that on the negative side its absolute necessity has been demonstrated and on the positive its use will give right results. There is now a considerable body of experience showing in what regions the quantum is required and in what the classical mechanics will prove sufficient. The principal fields in which the quantum theory has achieved success are (in historical order): (1) Planck's theory of radiation, (2) Einstein's theory of the photoelectric effect, (3) the work of Einstein, Debye and others on the specific heats of solids, (4) Bohr's theory of spectra, much the most remarkable of all. It has also been applied, with only moderate success, to the problem of the equation of state of gases. Here, while it explains the phenomena of specific heat quite well, the classical mechanics seem on the whole capable of dealing with the law of pressure. The kindred problems for which the classical mechanics appear to be adequate without the quantum are such as the diffraction, interference, dispersion, and anomalous refraction of light waves, the scattering of light by small particles and of X-rays by electrons, and such matters as the molecular theory of the viscosity of gases.

The essential fact of the quantum theory is the existence of a certain universal constant, the quantum, discovered by Planck and denoted by h. Its physical dimensions are mass $\times$ (length)^2 $\times$ (time)$^{-1}$, that is the same as energy multiplied by time (the "action" of general dynamics) or as angular momentum. Its magnitude is $6{\cdot}5 \times 10^{-27}$ erg. sec. The rule for its use may be stated generally as follows. If a system can vibrate with a frequency of ν vibrations per second, then the energy associated with this vibration can only be exchanged with other systems in quantities which are exact multiples of $h\nu$. Thus radiation of frequency ν will be absorbed by matter in amounts $h\nu$, and conversely if an electron radiates energy E the wave will have frequency E/h. This rule lacks precision and is inadequate for Bohr's spectrum theory (see below), but it covers all the other applications and they will show how it is to be interpreted. Of these the simplest is the photoelectric effect, and though it was not the first development in order of time, yet it will be convenient to treat of it here first.

§ (2) **Photoelectric Effect.**—In the photoelectric effect [1] light falls on a metallic surface and electrons are in consequence emitted from the surface. Let the frequency of the light waves be ν, then according to the quantum rule the electrons come off with energy $h\nu$. This is in fact observed to happen, though the effect is complicated by subsequent collisions which reduce the speed of the electrons, and above all by the presence of an electric field at the surface which prevents their egress or reduces their emergent velocity. Thus the energy of none of the electrons is observed to be greater than $h\nu - E_0$, and there is a minimum frequency E_0/h for the light beyond which none are obtained. The energy of the electrons is quite independent of the intensity of the source of light, which only affects their number.

An enormously exaggerated photoelectric effect is obtained from the X-rays, as these are electromagnetic waves with frequencies about 5000 times as great as those of visible light. In this case the electrons are emitted with a very high velocity and E_0 is insignificant. The converse process has also been observed,

[1] See "Photoelectricity," Vol. II.

electrons falling on a surface with energy E give rise to X-rays and none of these have a frequency greater than E/h. That lower frequencies are also observed is due to the fact that many of the electrons do not lose all their energy in a single step. In the hands of Millikan[1] this process has been made very successful for determining the value of h.

It is hardly necessary to insist on the inadequacy of classical mechanics to deal with the photoelectric effect. In the first place the effect would naturally be attributable to some sort of resonance, whereas its character is quite different. Also on the classical theory an electron should only be able to absorb energy which passed in its immediate neighbourhood, and it is easily calculated that for feeble sources of light this energy is so small that it would take years for the electron to accumulate the amount of energy which it in fact exhibits at once. Before the advent of the quantum theory attempts were made to overcome this difficulty by a quasi-corpuscular theory of light. Any such theory, however, encounters the insuperable objection that the same light which produces the photoelectric effect can also exhibit interference and all the other phenomena which connote that it is extended in the form of continuous waves. The combination of these two apparently mutually exclusive ideas of continuity and discontinuity is typical of the quantum theory.

§ (3) Partition of Energy.—For dealing with radiation and specific heats the idea of temperature must be introduced and associated with the quantum. In this connection the fundamental question is that of the partition of energy, which may be stated as follows. If a system of a large number of degrees of freedom can be separated into parts, what will be the average energy of each of those parts? It will be sufficient here to limit the generality to the case where the energy can be expressed as the sum of a set of squared terms. In this case classical mechanics gives the answer that every squared term takes on the average as much as every other, an amount $\frac{1}{2}kT$, so that a vibratory degree of freedom (for which the energy is half kinetic and half potential) has kT. Here T is the absolute temperature and k the "atomic gas constant," that is the ordinary gas constant divided by the number of molecules in a gramme-molecule. This is called the theorem of the Equipartition of Energy. But the quantum theory gives a very different law of partition and in the theory of radiation and of specific heats it is this other law that is important. According to this law, if there is a vibrator in the system which can only exchange energy in finite units and if ϵ be the value of this unit, then the average energy of this vibrator is not kT, but $\epsilon/(e^{\epsilon/kT}-1)$.

[1] See "Photoelectricity," § (3), Vol. II.

The formal proof of this expression is a rather complicated matter, depending as it does on difficult considerations of probability. It will suffice here to deduce it by analogy from the case of classical mechanics, where the corresponding theorem can be fully established. Suppose a system composed of a set of vibrators, each of any frequency, which are capable of exchanging their energy say by means of a few freely moving molecules, which can collide with them. Except for the few moments when collisions are occurring, each vibrator has a definite energy half kinetic and half potential, and every now and then this energy is altered by a collision. If the vibrator is observed for a very long time the energy will take on various values, but some more frequently and for longer than others. The probability that the vibrator has energy between E and $E+dE$ is defined as the fraction of the total time during which its energy lies between those limits. Then it is a demonstrable consequence of the laws of dynamics that this probability has the value $Ae^{-E/kT}dE$, where A is a constant that is readily determined to be $1/kT$. The average energy of the vibrator is therefore $\int_0^\infty E.Ae^{-E/kT}dE$. The value of this expression when integrated is kT, which is independent of E, and this is the theorem of the equipartition of energy. But if, as in the quantum theory, energy can only be taken up in finite units the formula for the probability must be altered, as E is no longer infinitely divisible, but can only be a multiple of ϵ. By analogy with the classical theory we therefore say that the probability of a vibrator with $n\epsilon$ of energy is $A'e^{-n\epsilon/kT}$, where A′ is a constant, provided that n is an integer, and is zero if it is not. By summing the probabilities of all possible arrangements A′ is seen to have the value $1/(1-e^{-\epsilon/kT})$ and the average energy is now found to be $\epsilon/(e^{\epsilon/kT}-1)$. This is the partition formula resulting from Planck's theory. If ϵ is small the expression reduces to kT the equipartition value, but for larger values of ϵ there is less energy. If ϵ is much larger than kT, the partition amount is approximately $\epsilon e^{-\epsilon/kT}$, which is very small compared with kT.

This result has been obtained by extending the form from the case where equipartition holds. An example will suffice to show the probability basis on which it really rests. Consider a system composed of four vibrators, of which two, A and B, can hold energy units ϵ, while the other two, C and D, require units 2ϵ, and suppose that the total energy of the system is 4ϵ. The rule for the application of prob-

ability is then that equal probability attaches to every possible different distribution of the energy, in which no distinction is to be made as to the identity of the units of energy. Thus if *a*, *b* and *c*, *d* denote respectively units ϵ on A, B and units 2ϵ on C, D, then all the possible arrangements are: *aaaa*, *bbbb*, *aaab*, *abbb*, *aabb*, *aac*, *aad*, *bbc*, *bbd*, *abc*, *abd*, *cc*, *dd*, *cd*, and each of these is to be counted as equally likely. Then the average energy on A or B is $\frac{1}{1}\frac{2}{3}\epsilon$ and on C or D is $\frac{1}{1}\frac{3}{3}\epsilon$. This illustrates the tendency for the energy to go into the vibrations which have the smaller units. The same process carried out with large numbers leads precisely to the above partition formula. That the probability basis is correct is confirmed by the fact that in mechanics energy is infinitely divisible, and if that is taken to be so in the above work, then the ordinary equipartition formula results.

Thus, since ϵ has the value $h\nu$, the consequence of the quantum theory is that if a vibratory system has frequency ν, and is at temperature T, its average energy is not kT, but

$$\frac{h\nu}{e^{h\nu/k\mathrm{T}}-1}.$$

§ (4) Radiation.—It was in his study of the radiation problem that Planck discovered the quantum, and it was in the same connection that Poincaré demonstrated rigorously the inadequacy of the older mechanics. A very considerable amount can be discovered about radiation by the application of thermodynamics, a branch of physics that must certainly be considered more fundamental than dynamics. It can be shown (see "Radiation," § (5), (ii.)) that there must be a universal formula describing the equilibrium of energy between radiation and matter in an enclosed space, and that this formula must be of the form

$$\mathrm{E}_\lambda=\frac{1}{\lambda^5}f(\lambda\mathrm{T})$$

(this is Wien's Displacement Law), where $\mathrm{E}_\lambda d\lambda$ is the energy (isotropic) per cubic centimetre in the spectral region between λ and $\lambda+d\lambda$, and T is the absolute temperature. All the thermodynamical conditions, however, are satisfied by any continuous function f, and so it is necessary to have recourse to other principles in order to find the form of f. In making use of the quantum to find the radiation formula it will be therefore allowable to choose any specially convenient system, and the apparent artificiality of the one chosen will be without influence on the result, for thermodynamics guarantee that this must also be the general formula.

The system chosen consists of a rectangular box with perfectly reflecting sides in which a small piece of matter is placed. This matter will radiate until the radiation in the box is in temperature equilibrium with it, that is until the radiation is the "complete" or "black" radiation corresponding to the temperature of the matter. If the matter is now withdrawn the radiation will be reflected backwards and forwards from the walls unchangeably, and can be easily analysed into a set of independent vibrations. The partition law can then be used to assign the average energy in each vibration.

Let the sides of the box be α, β, γ. Then the vibrations are described by means of the electric and magnetic forces, the typical expression being

$$\mathrm{X}=\Sigma\mathrm{A}_{lmn}\cos\frac{l\pi x}{\alpha}\sin\frac{m\pi y}{\beta}\sin\frac{n\pi z}{\gamma}\cos 2\pi\nu(t+\epsilon),$$

where l, m, n are integers, and ν is the frequency. By examining the electromagnetic equations it is found that

$$\frac{l^2}{\alpha^2}+\frac{m^2}{\beta^2}+\frac{n^2}{\gamma^2}=\frac{4\nu^2}{c^2}$$

(c is the velocity of light), and that there are two independent vibrations for each l, m, n corresponding to the two directions of polarisation. Now each vibration can be represented by marking the corresponding value l, m, n in a three-dimensional space. Then the number of vibrations for which ν is less than ν_0 is the same as twice the number of points marked inside the positive octant of the ellipsoid of semi-axes

$$\frac{2\alpha\nu_0}{c},\ \frac{2\beta\nu_0}{c},\ \frac{2\gamma\nu_0}{c}.$$

This number is

$$\frac{8\pi}{3}\alpha\beta\gamma\frac{\nu_0^3}{c^3}.$$

The number of vibrations with frequencies lying between ν and $\nu+d\nu$ is found by differentiation of this, and is

$$8\pi\,\alpha\beta\gamma\frac{\nu^2d\nu}{c^3}.$$

By the partition law each of these has average energy in amount $h\nu/(e^{h\nu/k\mathrm{T}}-1)$ and so the total energy in the box in frequencies between ν and $\nu+d\nu$ is

$$8\pi\,\alpha\beta\gamma\frac{\nu^2d\nu}{c^3}\frac{h\nu}{e^{h\nu/k\mathrm{T}}-1}.$$

If $\mathrm{U}_\nu d\nu$ is the energy per unit volume between ν and $\nu+d\nu$, then

$$\mathrm{U}_\nu=8\pi\frac{\nu^2}{c^3}\frac{h\nu}{e^{h\nu/k\mathrm{T}}-1}.$$

To change into the more usual form with wave-lengths instead of frequencies it is only necessary to apply the equations $\nu=c/\lambda$ and

$U_\nu d\nu = E_\lambda d\lambda$, and the result is Planck's radiation formula

$$E_\lambda = \frac{8\pi hc}{\lambda^5(e^{hc/k\mathrm{T}\lambda}-1)}.$$

If the equipartition law held, the formula of Rayleigh[1] would be given, viz.

$$E_\lambda = \frac{8\pi k\mathrm{T}}{\lambda^4}.$$

This formula is manifestly impossible as it implies infinite energy in the infinitely short wave-lengths, but it is inevitably deduced by every method that is based on classical mechanics. For the discussion of Planck's formula see "Radiation," § (6). Here it suffices to say that by determining the k of his formula from experiments Planck obtained one of the earliest really good values for the Avogadro constant. The values for h and k as determined by recent experiments are given by $h = 6{\cdot}55 \times 10^{-27}$ erg. sec. and $k = 1{\cdot}372 \times 10^{-16}$ erg./deg.

An attempt was made by Planck in his "Second Quantum Hypothesis" partially to reconcile quantum mechanics with ordinary mechanics by supposing that vibrators absorb energy continuously, but emit it in units. This theory leads to the same radiation formula as the other, but when applied in other directions it seems on the whole to raise more difficulties than it puts to rest. There is little doubt that no reconciliation in this way is possible, and most of the subsequent developments of quantum theory have proceeded on Planck's First Hypothesis, in which both emission and absorption are discontinuous processes.

§ (5) Specific Heats of Solids.—Before the advent of the quantum theory the problem of specific heats encountered great difficulty. For it is clear from the emission of spectra, etc., that there are at any rate a considerable number of degrees of internal freedom in an atom, and each of these should require its full share of energy on the equipartition principle. The quantum removes this difficulty in that, if the frequencies corresponding to these degrees are high, the partition rule will, at ordinary temperatures, require very little energy for them. Subject, then, to this new condition, the theory fits very well. Thus the specific heat at constant volume of a monatomic gas is simply due to the kinetic energy of translation of its atoms. For N atoms it will be $\frac{3}{2}k$N. For most diatomic gases two further degrees of freedom must be added, as the molecule may rotate about two axes perpendicular to the line joining its atoms; the specific heat is therefore $\frac{5}{2}k$N. In solids the atoms must be supposed to have three degrees of potential energy as well as three of kinetic, and so the specific heat for N atoms will now be $3k$N. This is the law of Dulong and Petit—that the atomic weight multiplied by the specific heat is the same for all the elements. Now there are certain well-known exceptions[2] to this law; diamond has a specific heat far below the value it should have. It was also discovered that all the other elements have much too low specific heat when the temperature is very low; and conversely that at a temperature of 600° C. diamond is nearly normal. The elucidation of these facts has been one of the successes of the quantum theory. It should be remarked that here we shall only treat of specific heat at constant volume. This often differs considerably from the observed quantity specific heat at constant pressure, but if the compressibility and thermal expansion are known the one can be deduced from the other by thermodynamics.

The first attack on the problem was due to Einstein. Rubens had shown that such solids as rock-salt have a very strong absorption band for light in the far infra-red. This is to be attributed to the vibrations of the atoms in the crystal and tells us their frequency ν, and so provides a means of applying the quantum partition rule. Instead of being $3k$T, the average energy of an atom should be $3h\nu/(e^{h\nu/k\mathrm{T}}-1)$ and so, by differentiating, the specific heat of N atoms will be

$$\frac{3(h\nu)^2 e^{h\nu/k\mathrm{T}}}{k\mathrm{T}^2(e^{h\nu/k\mathrm{T}}-1)^2}\mathrm{N}.$$

This formula represents the falling off of the specific heat at reduced temperatures, and vanishes at the absolute zero as it should, but at very low temperatures it gives values that are well below the experimental. An improvement was made by the empirical formula of Nernst and Lindemann, who replaced half the expression by a similar term involving $\nu/2$ instead of ν. This represents the experimental results much better, but still gives low values at the very low temperatures; moreover, there seems no theoretical reason for the presence of the octave terms.

The theory was greatly improved by Debye. The atoms of the solid are not to be regarded as having a single definite frequency, but will vibrate according to a set of normal modes that can be calculated from the theory of elasticity. Now if it were correct to suppose that the substance was a continuous elastic medium, it would follow by a method similar to that already used in the radiation problem that the number of degrees of freedom of vibration with frequencies between ν and $\nu + d\nu$ was

$$4\pi\left(\frac{1}{a_1^3}+\frac{2}{a_2^3}\right)\nu^2 d\nu,$$

where a_1 and a_2 are respectively the velocities of transmission of longitudinal and transverse waves. This number, of course, becomes infinite with ν, but it is obvious that there are really only a finite number of degrees of

[1] "Remarks upon the Law of Complete Radiation," *Phil. Mag.*, 1900, xlix. 539.

[2] See "Calorimetry, Quantum Theory," Vol. I.

freedom—3N for N atoms. Debye's method consists in supposing that all the frequencies below a certain value ν_m occur, and beyond these none. ν_m is determined by the condition that there shall be only 3N degrees of freedom in all, that is by the equation

$$\frac{4}{3}\pi V\left(\frac{1}{a_1^3}+\frac{2}{a_2^3}\right)\nu_m{}^3=3N,$$

where V is the volume containing the N atoms. Applying the quantum partition principle, the energy is given as

$$\int_0^{\nu_m} 4\pi V\left(\frac{1}{a_1^3}+\frac{2}{a_2^3}\right)\frac{h\nu}{e^{h\nu/kT}-1}\nu^2 d\nu$$

or

$$9N\int_0^{\nu_m}\frac{h\nu^3 d\nu}{(e^{h\nu/kT}-1)\nu_m{}^3}.$$

The differentiation of this expression by the temperature gives the specific heat at constant volume. If a "critical temperature" Θ is defined by the equation $k\Theta=h\nu_m$ (Θ may be calculated from a_1 and a_2, that is from the elastic constants), then the specific heat may be written as

$$3kf\left(\frac{\Theta}{T}\right)N,$$

where

$$f\left(\frac{\Theta}{T}\right)=3\left(\frac{T}{\Theta}\right)^3\int_0^{\frac{\Theta}{T}}\frac{x^4e^x}{(e^x-1)^2}dx.$$

When T is large f becomes equal to unity. For small temperatures, on the other hand, the specific heat becomes

$$3\,.\,k\,.\,77{\cdot}94\left(\frac{T}{\Theta}\right)^3 N.$$

This equation is verified fairly satisfactorily by experiment.

Debye's method of counting the vibrations is avowedly only rough, and his estimate is least accurate in precisely the place where the greatest accuracy in counting is needed, that is for the frequencies near ν_m. An improvement has been made by Born and Karman, who have succeeded in calculating the actual normal modes of vibration of a rock-salt crystal. They obtain the necessary constants from the elastic properties, and get values for the specific heat even better than Debye's. The actual comparison with experiment is described in the article in Vol. I. on "Calorimetry, Quantum Theory."

The fact that the specific heat at low temperatures is proportional to the cube of the temperature implies that the entropy near the absolute zero is a finite quantity; and by choice of the arbitrary constant which occurs in the definition of entropy it may be taken to be zero at the absolute zero. This is intimately related with Nernst's heat theorem, which has had such important consequences in physical chemistry.

§ (6) Specific Heat of Gases.—The application of the quantum to the specific heats of gases is not so satisfactory as for solids, partly because the experimental data are much less accurate, so that a discrimination between rival theories is not so easy. At both high and low temperatures the specific heat changes. At the high it increases—a certain part of the increase is attributable to dissociation, but it appears also that some of it is due to a participation in the energy by some of the internal degrees of freedom of the molecule. At low temperatures a very remarkable process occurs in the diatomic gases. Thus diatomic hydrogen at 40° abs. has all the characters of a monatomic gas—its specific heat at constant volume is $\frac{3}{2}k$N, and the ratio of the specific heats is $\frac{5}{3}$, instead of the ordinary values $\frac{5}{2}k$N with a ratio $\frac{7}{5}$.

The molecules of a diatomic gas can take energy in three ways. There is, first, the translational motion of the whole molecule; as would be expected, this method appears to be the same for all types of molecules. Secondly, there are the internal vibrations; the most important of these is the vibration in which the two atoms oscillate along the line joining them. It appears that for the permanent gases this vibration has too high a frequency for it to acquire any energy at all at ordinary temperatures, though it probably plays a part in the rise of specific heat at high temperatures. The third factor is the rotation of the molecules, and the treatment of degrees of freedom depending on rotation is a much more doubtful question. One of the most interesting methods is the following. Suppose a molecule to be rotating with a frequency ν, then its angular velocity is $2\pi\nu$. If I is its moment of inertia the kinetic energy is $\frac{1}{2}I(2\pi\nu)^2$. If the quantum principle is applicable this must be equal to $\frac{1}{2}nh\nu$, where n is an integer. (The reason for the factor $\frac{1}{2}$ will be seen in § (7) from the rule for the quantisation of orbits.) Now suppose that I is independent of ν, though there does not seem to be any cogent reason why this should be so in view of the varying centrifugal force. Then, since $\nu=nh/4\pi^2 I$, the molecule is only permitted to rotate with frequencies which are multiples of $h/4\pi^2 I$. There is evidence that frequencies of this type do exist from the presence of absorption bands in the far infra-red. The permissible energies of rotation will be $n^2h^2/8\pi^2 I$. Several methods have been tried for finding the specific heat from this. That which seems preferable theoretically is due to Ehrenfest. He works out the partition law *ab initio* in the same sort of way as that sketched in § (3), and it leads to the following formula for the total energy of N molecules:

$$\frac{e^{-\sigma}+4e^{-4\sigma}+9e^{-9\sigma}+\;.\;.\;.}{1+e^{-\sigma}+e^{-4\sigma}+e^{-9\sigma}+\;.\;.\;.}\sigma kTN+\frac{3}{2}kTN,$$

where $\sigma = h^2/8\pi^2 IkT$. The specific heat follows by differentiation. There are other theories, which give as good or nearly as good agreement with experiment—for instance, those of Bjerrum and of Krüger.

Various attempts have also been made to introduce the quantum in connection with other properties of gases. For instance, Tetrode and Keesom have examined the result of quantising the acoustic vibrations of a gas. This, of course, destroys the validity of the Maxwell law of distribution of velocities, and the equation of state is modified. The attempt is not very successful, and the resulting equation differs from the actual at least as much as that obtained from classical mechanics.

Mention should also be made of some theories of Planck, Sackur, and Tetrode, who have tried to introduce the quantum in a still more fundamental way by using it to define the probability basis which controls the distribution of molecules. The object was to calculate the condensation point of a gas, but none of the theories have been very successful.

§ (7) Spectrum Theory.—It will suffice merely to indicate the general outline of the Bohr's theory of spectra, which is the most remarkable application of quanta yet made. The spectra of many of the elements have been analysed into fairly simple formulae, and these formulae are completely different from those found in the ordinary theory of vibrations. For the frequency of each line is given by taking the difference of a pair of *terms*, and the separated terms usually fall into an algebraic series of some kind. This rule for constructing lines from terms is called the Combination Law of Ritz. For example, the series-terms for the primary hydrogen spectrum are of the form N/n^2, where N is a constant and n any integer, and so the lines are

$$N\left(\frac{1}{n^2} - \frac{1}{n'^2}\right).$$

The famous Balmer[1] series is the particular case where $n=2$, while n' has all integral values greater than 2. Spectrum work is usually carried out not in frequencies, but in reciprocal wave-lengths; the constant N/c is called Rydberg's constant. Now in dealing with the photoelectric effect it was seen that a loss of energy E was associated with radiation of frequency E/h. Therefore if an atom of initial energy E_1 radiates, and in consequence has its energy reduced to E_2, the radiation will be of frequency $(E_1 - E_2)/h$. Thus the quantum theory directly explains the meaning of the combination law; if any series-term is multiplied by h the result is the energy either before or after the emission.

[1] See also article "Spectroscopy, Modern," Vol. V.

The reverse argument applies for the absorption of radiation. Thus if an atom is in the arrangement corresponding to energy E_2, and is irradiated with radiation of frequency $(E_1 - E_2)/h$, it may absorb it and thereby rise to the arrangement with energy E_1.

This use of the quantum is only half the solution, however; for it is still necessary to find the arrangements corresponding to the series *terms*, and the laws of ordinary mechanics will not determine them definitely. For example, the hydrogen atom consists of a nucleus and one electron, which describes an elliptic orbit about it, and this it may do with any radius and any energy whatever. So in order to make the energy of the atom definite a second use of the quantum must be made. This consists in "quantising the orbits," which means that some of the orbital constants are determined by the quantum, so that of the infinity of orbits possible on dynamical principles only a certain limited series is permissible. For example, consider first only the circular orbits which the electron may describe. The electron can absorb or emit energy in multiples of $h\nu$; in addition to this the further condition is assumed that the angular momentum is to be a multiple of $h/2\pi$. This is a consequence of the rule of quantisation applied to this case. Thus if a is the radius and v the velocity of the electron, the motion is given by

$$\frac{mv^2}{a} = \frac{e^2}{a^2},$$

where e is the electric charge of electron and nucleus, and m is the mass of the electron. The quantisation gives $mva = nh/2\pi$. From these equations it follows that

$$a = \frac{n^2h^2}{4\pi^2e^2m} = n^2a_1,$$

and the energy is

$$E_n = -\frac{2\pi^2e^4m}{h^2}\frac{1}{n^2}.$$

Thus, if the electron shifts from the orbit of radius n^2a_1 to that of radius n'^2a_1 there will be an emission of the line defined by

$$\nu = \frac{2\pi^2e^4m}{h^3}\left(\frac{1}{n'^2} - \frac{1}{n^2}\right).$$

This is precisely Balmer's expression, and the agreement is numerical. Thus the Rydberg constant for hydrogen is directly observed to be 109677·691 cm.$^{-1}$; if the most accurate known values of e, m, h, and c are substituted in the above expression the result is $1{\cdot}095 \times 10^5$ cm.$^{-1}$. The value of a_1, the least permissible radius, is about $0{\cdot}5 \times 10^{-8}$ cm., which is of the same order as the known size of atoms.

This is the substance of Bohr's original discovery. The second great advance is the extension of the principle by Sommerfeld to

the case of more complicated orbits. In these cases the principle of quantisation can best be expressed in the language of general dynamics. Thus in a periodic or quasi-periodic motion let q be one of the co-ordinates which define the motion, and p the corresponding momentum; then q and p will fluctuate about certain values during the period. The relation which expresses the quantisation of the orbit is $nh=\int pdq$, where the integral is taken round one period, and this is to be applied for each of the variables q. In the previous example q was the vectorial angle and ranged from 0 to 2π, and p, the angular momentum, is therefore a multiple of $h/2\pi$.

In elliptic motion the quantisation is naturally more complicated, as it applies to both θ and r. The angular momentum is again constant, so that the quantisation gives $p_\theta=n_1h/2\pi$. The radial momentum is

$$m\frac{dr}{dt} \quad \text{or} \quad \frac{p_\theta}{r^2}\frac{dr}{d\theta},$$

and the quantum relation for this is

$$n_2h=\int p_r dr=\int\frac{p_\theta}{r^2}\left(\frac{dr}{d\theta}\right)^2 d\theta,$$

where the integration is to be taken from the minimum of r to its maximum and back again, that is once round the ellipse. Now if a is the half major axis and ϵ the eccentricity, the equation to the ellipse is $a(1-\epsilon^2)/r=1+\epsilon\cos\theta$ and p_θ is given by the relation $p_\theta{}^2=e^2ma(1-\epsilon^2)$. Substituting for r in terms of θ and performing the integration we have $n_2h=p_\theta\,.\,2\pi[(1-\epsilon^2)^{-\frac{1}{2}}-1]$. Hence the permissible orbits are those for which

$$\frac{n_1}{n_1+n_2}=\sqrt{1-\epsilon^2}$$

and $$a=\frac{(n_1+n_2)^2h^2}{4\pi^2e^2m}=(n_1+n_2)^2a_1.$$

The total energy is found to be

$$E_{n_1,\,n_2}=-\frac{2\pi^2e^4m}{h^2}\frac{1}{(n_1+n_2)^2}$$

As n_1 and n_2 are integers, each permissible orbit has major axis equal to the diameter of one of the circular orbits obtained above, and this diameter determines the energy, but the eccentricity can only have certain prescribed values. For example, if $n_1+n_2=3$, we have three, and only three, different orbits: (1) $n_1=3$, $n_2=0$, a circle of radius $9a_1$; (2) $n_1=2$, $n_2=1$, an ellipse of which the semi-axes are $9a_1$, $6a_1$, and the eccentricity is $\sqrt{5}/3$; (3) $n_1=1$, $n_2=2$, an ellipse of semi-axes $9a_1$, $3a_1$, and eccentricity $\sqrt{8}/3$. The case $n_1=0$, $n_2=3$ is supposed to be inadmissible, as it would involve a collision of the electron with the nucleus. The red hydrogen line $N(1/2^2-1/3^2)$ is thus made up of six components derived from the change from any one of these three orbits to either of the orbits $n_1'=2$, $n_2'=0$, or $n_1'=1$, $n_2'=1$.

Sommerfeld also took into account the fact that the mass of the electron is a function of its velocity. This modifies the orbits in such a way that the ellipses have slightly different energies. The consequence is that each line of the Balmer series has a fine structure of a type he was able to predict, which has been confirmed by experiment. The method has also been applied, with complete success, to the influence on spectra of electric and magnetic fields, the Stark and Zeeman effects. For these cases, in order to predict the polarisation of the component lines, it is necessary to introduce supplementary principles. Sommerfeld has suggested a "principle of selection" (*Auswahlprinzip*) depending on the idea of the momentum of radiation. In addition to giving a polarisation rule, it has the consequence that certain lines, which would on the original theory have been expected, will not be found, as is experimentally the case. An alternative, which appears preferable, is Bohr's "principle of correspondence." It is not possible to describe this here, but the general idea is that the behaviour of a quantised system may be predicted by assimilating the result which would be obtained when the number of quanta is large, to the result given on the principles of classical mechanics.

As points of further interest in the theory, mention may be made of the fact that ionised helium gives a spectrum precisely the same as hydrogen, except that the constant N is changed. Here, as in hydrogen, there is a single electron moving round a nucleus, but now the nuclear charge is $2e$. Hence the permissible orbits are all half the size of the others, and the corresponding energies are four times as great. The new N is almost but not quite exactly four times as great as the old, and the departure from exactness has been explained very satisfactorily. It is attributable to the fact that the mass of the nucleus is not infinite compared to that of the electron, so that the nucleus itself has a small motion. As the helium nucleus is four times as heavy as the hydrogen, there will be a difference in the two motions, with a resulting effect on Rydberg's constant. It is possible in this way to evaluate the ratio of the mass of an electron to that of a hydrogen atom from purely spectroscopic data. Another point of interest is that the X-ray spectra can be analysed into terms, many of which fit into series of exactly the same form as those of hydrogen and helium, and these lines exhibit the same fine structure too, but that the separation of the components is much greater. A very considerable theory of these spectra exists.

By using (α) Rydberg's constant, (β) the difference of this constant for hydrogen and helium, (γ) the "constant of fine structure" and the value of the velocity of light, it is possible to work out from spectroscopic data alone all the fundamental elementary constants e, m, and h. Of the three relations (α) is known so accurately that the precision depends entirely on the other two. Of these (β) determines what is in effect e/m in terms of the small difference between two large, but accurately known, numbers; its precision is about equal to that from direct methods. (γ) gives a relation between e and h in terms of the separation of the lines in the fine structure, and the observations are exceedingly difficult. The resulting values from the three relations are nearly as good as those by other methods.

TABLE OF VALUES OF THE QUANTUM

$h \times 10^{27}$.	Method.	Notes.
$6{\cdot}551 \pm 0{\cdot}009$	Total radiation	A
$6{\cdot}557 \pm 0{\cdot}013$	Spectral radiation	B
$6{\cdot}578 \pm 0{\cdot}026$	Photoelectric effect	C
$6{\cdot}555 \pm 0{\cdot}009$	X-rays	D
$6{\cdot}579 \pm 0{\cdot}021$	Ionisation potential	E
$6{\cdot}526 \pm 0{\cdot}200$	Spectroscopic data	F
$6{\cdot}547 \pm 0{\cdot}009$	Semi-spectroscopic data	G

A is determined by taking k known in the formula for the total radiation.

B is determined from the position of the maximum intensity of black radiation at any temperature.

C is found directly from observations on the photoelectric effect.

D is determined from the minimum wave-length of X-rays excited by electrons of known energy.

E depends on observations that are half spectroscopic. From spectroscopy the energy of an atom may be known. The ionisation potential gives another measure of this, as it shows the energy necessary for the removal of an electron.

F is determined by the purely spectroscopic method involving the above relations α, β, γ.

G is determined by the use of (α) and (β) as in F, but (γ) is replaced by using the known value of e.

Of these methods all except F depend on assuming Millikan's value $4{\cdot}774 \times 10^{-10} \pm 0{\cdot}009$ E.S.U. for e, the electronic charge; but the close agreement between A and B is a direct support for that value.

C. G. D.

BIBLIOGRAPHY

For most of the quantum theory it is unnecessary to refer to the original papers, as they have been incorporated in many books. Among others the following are works that may usefully be consulted.

La Théorie du rayonnement et des quanta (Gauthier Villars, Paris, 1912) is an account of a congress held at Brussels in 1910 of leading physicists. All the earlier aspects of the quantum theory are very fully discussed.

Report on Radiation and the Quantum Theory, by J. H. Jeans (Lond. Phys. Soc., 1914), is a concise account of the whole field of the quantum theory, except that spectrum theory has developed a great deal further since its publication.

A System of Physical Chemistry, vol. iii., by W. C. McC. Lewis (Longmans, Green & Co., 1919), gives a very detailed, but uncritical, account of many parts of the quantum theory, without the use of much mathematics.

Atombau und Spektrallinien, by A. Sommerfeld (Vieweg, 1919), gives a very complete account of atomic physics, with special reference to the Bohr theory of spectra.

The Dynamical Theory of Gases, by J. H. Jeans, Third Edition (Camb. Univ. Press, 1920), contains a few chapters dealing mathematically with the chief points of the theory.

Report on the Quantum Theory of Spectra, by L. Silberstein (Adam Hilger, 1920), is a short but complete account of spectrum theory.

References to original papers will be found in these works. Bohr's "Correspondence Principle" is dealt with in two papers "On the Quantum Theory of Line Spectra" by N. Bohr (*D. Kgl. Dansk. Vidensk. Selsk. Skr.*, 1918); the papers are in English.

QUARTZ, anomalous dispersion of. See "Wave-lengths, The Measurement of," § (7).

Crystalline form of. See "Quartz, Optical Rotatory Power of," § (2).

Optical rotatory power of, for some important wave-lengths, tabulated. See *ibid.* § (4), Table II.

Optical systems, used in microscopy with ultra-violet light. See "Microscopy with Ultra-violet Light," § (3).

QUARTZ, OPTICAL ROTATORY POWER OF

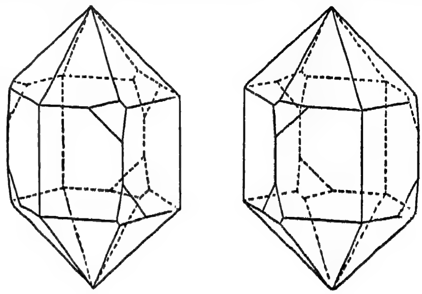
§ (1) HISTORICAL.—The phenomenon of optical rotatory[1] power was discovered by Biot, who communicated to the Institute of France on November 30, 1812, a memoir of 371 pages, "On a New Kind of Oscillation which the Molecules of Light experience in traversing certain Crystals." Arago, in 1811,[2] had found that a plate of quartz, interposed between a polariser and analyser, was capable of depolarising the light in such a way that transmission took place where previously there had been complete extinction. The transmitted light was not white but coloured. Thus with increasing thickness of the plate the colours changed progressively through the series—yellow, orange, rose-red, violet, blue, and green. These colours were shown by Biot to be due to a rotation of the plane of polarisation, which increased (i.) with the thickness of the plate, (ii.) with change of colour from red to violet. It was therefore impossible, when a beam of polarised light had passed through a quartz plate, to extinguish all the colours simultaneously. The tints which Arago observed were due to the selective extinction of light which had been rotated through 180° (or a multiple of 180°) by the plate of quartz. Thus a plate about

[1] See also "Polarised Light," § (20), and "Polarimetry," § (1).

[2] *Mem. Inst.*, 1811, pp. 93-134.

4 mm. thick which rotates the plane of polarisation of violet light through 180° gives rise to a yellow tint. As the thickness of the plate increases the absorption band moves towards the red end of the spectrum, but before it has disappeared into the infra-red a second absorption band moves in from the ultra-violet, so that a plate of quartz about 9 or 10 mm. thick gives an extinction both at the red end of the spectrum, where the plane of polarisation is rotated through 180°, and in the blue or violet region, where it is rotated through 360°. These extinctions at the two ends of the spectrum give rise to the green colour, which is the last member of the series of colours recorded above. Thicker plates give extinctions which increase progressively in number and in the closeness with which they are packed in the spectrum. No marked coloration is then observed in the transmitted light, although the absorption bands can be seen very clearly with the aid of a spectroscope. A quartz plate, 3·65 mm. in thickness, produces a rotation of 90° in the bright orange-green region of the spectrum, a little on the yellow side of the green mercury line of wave-length 0·5461 μ, which gives a rotation of 93° in a plate of this thickness. When the polariser and analyser between which the quartz plate is inserted are parallel instead of crossed the brightest part of the spectrum is extinguished and a rose-violet colour, known as the *neutral tint*, is produced which is very sensitive to small changes in the setting of the polariser or analyser, or in the thickness of the plate. This fact was formerly utilised in the construction of *polarimeters* for measuring the rotation of polarised light in various media.

§ (2) Crystalline Form.—The typical crystals of quartz have the form of a hexagonal prism capped by two hexagonal pyramids, and are therefore referred to the "hexagonal system" in which the faces are located most conveniently by their intercepts on a vertical principal axis and three horizontal axes inclined at 60° to one another. A closer examination shows the presence on some crystals of small facets (*Fig.* 1, (*a*) and (*b*)) which reduce the symmetry of the crystal by destroying the six vertical planes of symmetry intersecting in the principal axis, which are present in the simple hexagonal prism. Even when these facets are absent the lower symmetry of the quartz crystals can be shown by studying the tiny "etched-figures" obtained by the action of hydrofluoric acid on the crystals. When these factors are taken into account it is seen that the quartz crystals are not "holohedral" but "tetartohedral," the smaller facets including only six facets of each form instead of twenty-four as in the holohedral system of symmetry, where each face of the hexagonal prism would carry a facet on each corner. The symmetry of the quartz crystal is in fact restricted to one vertical axis of threefold symmetry, and



(*a*) (*b*)

Fig. 1.—(*a*) Dextrorotatory Quartz; (*b*) Laevorotatory Quartz.

three horizontal axes of twofold symmetry; these axes of symmetry coinciding with the four crystallographic axes used in locating the faces in the hexagonal system. Since no centre of symmetry and no plane of symmetry is present, the crystals can exist in two enantiomorphous forms as in *Fig.* 1, (*a*) and (*b*). The optical rotatory power of quartz is observed in plates cut perpendicularly to the principal axis, and these two types of crystals rotate the plane of polarisation of light to an equal extent but in opposite directions; twin crystals containing both forms of quartz are also found. Herschel, who discovered the relationship between the crystal form and rotatory power of quartz, reversed Biot's convention by describing as dextrorotatory a plate of quartz which turns the plane of polarisation to the right *as viewed from the polariser, i.e.* in the opposite direction to dextrorotatory camphor or sugar, which rotate the plane of polarisation to the right *as viewed from the analyser*, since in the case of organic compounds Biot's convention is always followed. Herschel's convention is still used by opticians, but in *Fig.* 1 Biot's convention is used.[1]

§ (3) Rotatory Dispersion. (i.) *Biot's Law.*—Biot's investigations revealed the existence of two types of *optical rotatory dispersion*, as indicated in the following data:[2]

	Quartz.	Tartaric Acid.
Red	18·00°	38° 7′
Yellow . . .	21·40°	40° 20′
Orange . . .	23·99°	42° 51′
Green . . .	27·86°	46° 11′
Blue. . . .	32·31°	44° 40′
Indigo . . .	36·13°	42° 9′
Violet . . .	40·88°	39° 38′

[1] *Nature*, 1922, cx. 808.
[2] *Mém. Acad. Sci.*, 1838, xv. 236.

In the case of quartz the rotations increased progressively as the wave-length diminished, and Biot concluded that "the rotation of the different simple rays is reciprocal to the square of their wave-lengths"; [1] in the case of tartaric acid the rotations passed through a maximum in the green. Turpentine and cane-sugar appeared, however, to be identical with quartz in their rotatory dispersion, since the optical rotations which they produced could be neutralised completely with the help of a quartz plate of opposite sign and suitable thickness acting as a *compensator*; this fact has been utilised in the construction of *saccharimeters*, in which the strength of a solution of sugar is determined by balancing its optical rotation with that produced by a variable thickness of quartz.

(ii.) *Drude's Theory.*—Biot's law, $\alpha=k/\lambda^2$, was proved by more exact measurements to give only an approximate representation of the variations with wave-length of the rotatory power of quartz, cane-sugar, etc., and was modified by Boltzmann, whose equation may be expressed as $\alpha=k_1/\lambda^2+k_2/\lambda^4$. This equation was purely empirical and did not provide an adequate expression of the deviations from Biot's law; Drude, in his *Theory of Optics* (1900), put forward an equation based upon the electronic theory of radiation which accounts for the observed facts in a satisfactory manner. Drude's equation for light which does not approximate in frequency to that of the electrons of the optically-active medium takes the form

$$\alpha=\frac{k_1}{(\lambda^2-\lambda_1^2)}+\frac{k_2}{(\lambda^2-\lambda_2^2)}\ \cdots\ +\frac{k_n}{(\lambda^2-\lambda_n^2)},$$

where $\lambda_1^2, \lambda_2^2, \ldots \lambda_n^2$, are the wave-lengths corresponding with the various "free-periods" of the electrons of the medium, and k_1, k_2, k_3 are constants.

§ (4) EXPERIMENTAL VERIFICATION.—The writer has shown that a single term of Drude's equation provides an adequate expression both of the optical and of the magnetic rotatory dispersion in a large range of organic compounds,[2] whilst the phenomena of anomalous rotatory dispersion as observed, for instance, in the esters of tartaric acid are covered completely by making use of two terms of opposite sign.[3] In the case of quartz, a single term of Drude's equation serves to express the rotations to about 0·040° per mm. over a considerable range of the spectrum, as shown in Table I., where $\alpha_1=k(\lambda^2-\lambda_0^2)$, $k=7{\cdot}1764$, $\lambda_0^2=0{\cdot}01677$;[4] but Drude himself used two terms in order to cover the data given by Gumlich [5] for wave-lengths ranging from 2·140 μ to 0·21935 μ. In this equation he assumed the free periods, as deduced from measurements of refractive dispersion, etc., to be given by

$$\lambda_1^2=0{\cdot}010627,\quad \lambda_2^2=78{\cdot}22,\quad \lambda_3^2=430{\cdot}6.$$

It appeared, however, that the two free periods corresponding to radiations in the infra-red did not affect the rotations in the range covered by the observations of Gumlich, so that $k_2=k_3=0$. On the other hand, Drude found it necessary to introduce a term k'/λ^2 to express the fact that the rotations appeared to be influenced by ions "whose natural periods are extremely small, much smaller than those corresponding to λ_1," so that no constant corresponding to λ_1^2 or λ_2^2 need be introduced into this term of the equation. The equation thus deduced took the form

$$\alpha_2=\frac{k_1}{(\lambda^2-\lambda_1^2)}+\frac{k'}{\lambda^2},$$

where $\lambda_1^2=0{\cdot}010627, k_1=12{\cdot}200$, and $k_2=-5{\cdot}046$.

In the series of eighteen rotations quoted from Gumlich, which ranged from 1·60° to 220·72° per mm., the equation gave a maximum deviation of 0·28°/mm., and an average deviation of 0·06°/mm. between the observed and calculated values as set out in Table I.

TABLE I

DRUDE'S FORMULA FOR THE ROTATORY POWER OF QUARTZ

λ.	α (obs.).	One Term. α_2(calc.)	$100(\alpha-\alpha_2)$.	Two Terms. α_1(calc.)	$100(\alpha-\alpha_1)$
2·140μ	1·60	1·57	+3	1·57	+3
1·770	2·28	2·30	−2	2·29	−1
1·450	3·43	3·44	−1	3·43	±
1·080	6·18	6·24	−6	6·23	−5
0·67082	16·54	16·54	±	16·56	−2
0·65631	17·31	17·33	−2	17·33	−2
0·58932	21·72	21 71	+1	21·70	+2
0·57905	22·55	22·53	+2	22·53	+2
0·57695	22·72	22 70	+2	22·70	+2
0·54610	25·53	25·50	+3	25·51	+2
0·50861	29·72	29·67	+5	29·67	+5
0·49164	31·97	31·91	+6	31·92	+5
0·48001	33·67	33·59	+8	33·60	+7
0·43586	41·55	41·43	+12	41·46	+9
0·40468	48·93	48·85	+8	48·85	+8
0·34406	70·59	70·62	−3	70·61	−2
0·27467	121·06	..		121·34	−28
0·21935	220·72	..		220·57	+15

A series of readings to six significant figures of the rotatory power of quartz, made with the help of a column of dextro-quartz 181 mm. in length, and a column of laevo-quartz 226 mm. in length,[6] showed that the two-term formula employed by

[1] *Mém. Acad. Sci.*, 1817, ii. 49, 57, and 135; paper read Sept. 22, 1818.
[2] *Trans. Chem. Soc.*, 1913, ciii. 1067, and 1919, cxv. 300.
[3] *Ibid.*, 1915, cvii. 1187.
[4] Rupe und Akermann, *Annalen der Chemie*, 1920 ccccxx. 11; compare Lowry and Dickson, *Trans. Chem. Soc.*, 1913, ciii. 1072.
[5] *Wied. Ann.*, 1898, lxiv. 349.
[6] Lowry, *Roy. Soc. Phil. Trans.*, 1912, A, ccxii. 261-297.

Drude was far from exact even over the narrow range from Li 0·6708 to Hg 0·4358, within which visual readings can be taken. A satisfactory agreement was, however, obtained by reintroducing one of Drude's infra-red terms as in the equation

$$a_3 = \frac{k}{\lambda^2} + \frac{k_1}{\lambda^2 - \lambda_1^2} + \frac{k_2}{\lambda^2 - \lambda_2^2},$$

where $\lambda_1^2 = 0{\cdot}010627,$

$\lambda_2^2 = 78{\cdot}22,$

$k_1 = 11{\cdot}6064,$

$k_2 = 13{\cdot}42,$

$k = -4{\cdot}3685.$

This formula, in the case of twenty-two out of twenty-four wave-lengths, gave an average deviation between the observed and calculated rotations of only 0·001° per mm. or 1 part in 25,000, and this deviation has since then been reduced to one-half of its former magnitude by introducing for the wave-lengths the more exact values deduced in recent years from measurements with the interferometer; but even this formula becomes inexact when applied to the most exact measurements in the ultra-violet region of the spectrum. These measurements, made by a photographic method, with a column of laevo-quartz nearly half a metre in length, include some 700 wave-lengths, and can be expressed by the formula

$$a_4 = k + \frac{k_1}{\lambda^2 - \lambda_1^2} + \frac{k_2}{\lambda^2 - \lambda_2^2},$$

where

$\lambda_1^2 = 0{\cdot}012742, \quad \lambda_2^2 = 0{\cdot}000974,$

$k_1 = 9{\cdot}5644, \quad k_2 = -2{\cdot}3114, \quad k = -0{\cdot}1915.$

In this formula it has been necessary to introduce two "ultra-violet" terms, with dispersion-constants, λ_1^2 and λ_2^2, deduced independently of the constant 0·010627 which Drude had derived from measurements of refractive dispersion, and which is perhaps a sort of "mean" of the two constants set out above. The increased importance which is thus attached to the more distant ultra-violet absorption-band is counterbalanced by a diminution in the influence assigned to the infra-red bands; this is so slight that it is a matter of indifference whether it is expressed by a term $k/(\lambda^2 - 78{\cdot}22)$ or by a mere constant as in the equation now adopted, where the infra-red terms contribute only −0·1915° per mm. to a rotation which ranges from 16·5359° per mm. at Li 0·6707846 μ to 41·5487° at Hg 0·4358342 μ, and 187·225° at Fe 0·2327468 μ. This formula represents the rotations in this range with an average error of 0·0007° per mm. (or about 0·003 per cent in visual readings) for twenty-three wave-lengths, and 0·003° per mm. (or about the same percentage as for the visual readings) for 700 wave-lengths read by the photographic method. The problem of redetermining the rotatory power of quartz in the infra-red region has not been undertaken seriously, but the formula given above shows an average deviation of about 0·05° per mm. in comparison with the first series of new infra-red readings.

The optical rotatory power of quartz for some important wave-lengths is set out in Table II.

TABLE II

OPTICAL ROTATORY POWER OF QUARTZ

λ.	α.	$(\alpha - a_4)10^4$.
Li 0·6707844 μ	16·5359 °	−4
Cd 0·64384696	18·0225	−11
Zn 0·6362344	18·4786	−18
Na 0·5895930	21·7001	−13
Na 0·5889963	21·7483	+3
Hg 0·5790659	22·5455	+1
Cu 0·5782158	22·6157	+
Hg 0·5769598	22·7201	+
Cu 0·5700248	23·3101	−3
Ag 0·5471551	25·4318	+15
Ag 0·5465489	25·4911	+7
Hg 0·5460741	25·5371	−5
Tl 0·535065	26·6718	+6
Ag 0·5209084	28·2447	+4
Cu 0·5153226	28·9036	−2
Cu 0·5105547	29·4851	−7
Cd 0·5085822	29·7308	−10
Zn 0·4810534	33·5154	−15
Cd 0·4799922	33·6761	−12
Zn 0·4722162	34·8875	−13
Zn 0·4680138	35·5712	−3
Cd 0·4678163	35·6043	+
Hg 0·4358342	41·5487	10
λ.	α.	$(\alpha - a_4)10^3$.
Fe 0·5371493 μ	26·448 °	−3
Fe 0·5232958	27·966	−3
Fe 0·5001880	30·813	−1
Fe 0·4789657	33·830	−3
Fe 0·4647437	36·114	−4
Fe 0·4531155	38·162	−2
Fe 0·4375934	41·183	−1
Fe 0·4233615	44·289	+1
Fe 0·4118552	47·068	+
Fe 0·3935818	52·066	−4
Fe 0·3805346	56·163	+3
Fe 0·3640391	62·085	−1
Fe 0·3485344	68·586	−2
Fe 0·3323739	76·551	−14
Fe 0·3125661	88·483	−10
Fe 0·2941347	102·434	+
Fe 0·2813290	114·277	+4
Fe 0·2679065	129·200	+17
Fe 0·2413310	160·661	+2
Fe 0·2327468	187·225	+4

In this table the rotations are given to 0·0001°/mm. for twenty-three wave-lengths,

and the deviations from the formula are given to the fourth decimal place; rotations are also given to 0·001°/mm. for twenty of the eighty iron lines of the standard series, the deviations from the formula being given to the third decimal. In the case of other series of lines the wave-lengths are usually too uncertain to give any satisfactory agreement between the observed and calculated values.

T. M. L.

— R —

Radiation, Application of Quantum Theory to. See "Quantum Theory," § (4).

RADIATION, DETERMINATION OF THE CONSTANTS AND VERIFICATION OF THE LAWS

I. Introductory Discussion

§ (1) Statement of the Laws of Radiation.—The derivation of the laws of radiation from a uniformly heated enclosure or so-called black body are given elsewhere in this Dictionary.[1] It is therefore sufficient to say that the law of total radiation, generally called the Stefan-Boltzmann (2) law, states that the total emission of thermal radiation R of all wave-lengths passing from a uniformly heated enclosure at temperature T_1 to another body at a temperature T_0 is proportional to the difference of the fourth powers of their absolute temperatures, or

$$R = \sigma(T_1^4 - T_0^4). \quad . \quad . \quad . \quad (1)$$

In this formula σ is the coefficient or constant of total radiation under discussion. The numerical value of this coefficient is of the order $\sigma = 5{\cdot}7 \times 10^{-5}$ erg per cm.^{2} per sec. per deg.^{4}. In this paper only the first three significant figures will be mentioned, *e.g.* $\sigma = 5{\cdot}72$.

That the total radiation from a properly constructed and uniformly heated cavity is proportional to the fourth power of the absolute temperature is amply demonstrated within the experimental errors of observation and the accuracy of the temperature scale.

The law of spectral radiation, which specifies the distribution of thermal emission intensities in the spectrum of the radiation emitted by a uniformly heated enclosure, or so-called black body, is best represented by Planck's (3) formula

$$E_\lambda = c_1\lambda^{-5}(e^{C_2/\lambda T} - 1)^{-1}. \quad . \quad . \quad (2)$$

In this formula the constant C_2 determines, to a great extent, the slope of the isothermal spectral energy curve. Suffice it to say that these laws are based on well-grounded theoretical foundations, and after almost two decades of discussion they remain unchanged.

The object of this paper is to examine into the various instruments and methods employed in the determination of these two constants, σ and C_2, and to give an estimate of the results therewith obtained.

At first glance the various determinations of these constants appear to cover a wide range of numerical values. This is especially true of the values of the coefficient of total radiation, σ, in the measurement of which all sorts of ill-considered methods have been used. Moreover, no corrections were made for atmospheric absorption of the thermal radiation in its passage from the radiator to the receiver. As will be shown presently, most of these older determinations are in remarkably close agreement with the newer ones, when obvious corrections for atmospheric absorption and for reflection of the incident radiation from the receiver are introduced.

While no doubt it is desirable to conform to strict specifications and methods of procedure, in the ultimate result compromises have to be made. For example, it is conceded to be desirable to use a black body receiver as well as radiator, but most of the receivers of this type, thus far used, were usually of great heat capacity, hence slow acting and subject to errors; whereas the question of speed is an important factor in making measurements. Each method has some defect, but in some cases this is negligible in comparison with the accidental errors of experimentation.

The evaluation of radiant energy in absolute measure is accomplished by substitution methods; and the fault to be found with the various methods thus far employed is that they are unsymmetrical in their application. What is needed is some sort of quick-acting calorimeter in which the energy recovered can be compared with the energy supplied. But one can hardly expect to make accurate measurements on a radiator heated to say 1000° C. with an instrument that is adapted to measure the solar constant.

It may be noted that in many of the experiments which will now be described the radiators were operated at temperatures which were too low to eliminate properly the correction for temperature (radiation) of the shutters, diaphragms, etc., and also for the losses by air conduction in the receiver. On the other hand, some experimentors used radiators operated at a high temperature,

[1] See "Quantum Theory," § (4); "Radiation Theory," § (6); "Pyrometry, Optical," § (1), Vol. I.

thus obviating these difficulties; but then the receiver was a sluggishly acting instrument which could not be properly calibrated. To Féry is due the credit for abandoning the idea of using a radiator heated to 100° C., and adopting a radiator heated to 1000° C. or higher. In this manner the radiation to be measured is far in excess of the small disturbing factors, such as radiation from shutters, diaphragms, etc.

§ (2) Discussion of Black Body Radiation and Radiators.—In order to arrive at an intelligent understanding of what has been accomplished in the determination of the constants of radiation, and in order to indicate the way to future progress in this subject, it is important to consider some of the conceptions, as well as some of the most recent instruments and methods used in the production and measurement of the radiation emanating from a uniformly heated enclosure or so-called black body.

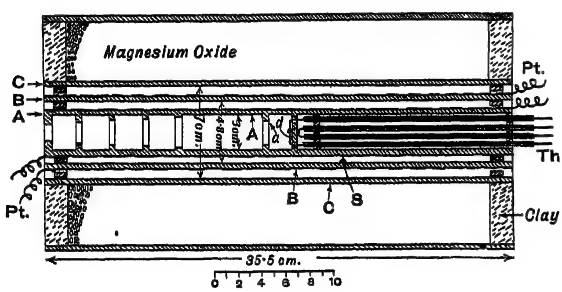


Fig. 1.

Every substance has a characteristic emission spectrum. On the other hand, the emission spectrum of a uniformly heated cavity is independent of the composition of the material forming the walls of the enclosure. It is a function only of the temperature.

The emission of thermal radiation from, and the absorption of thermal radiant energy of all wave-lengths entering into the cavity is complete. In a material substance this can occur only on the short wave-length side of a band of anomalous dispersion where the absorptivity is high but where the refractive index and hence the reflecting power is practically *nil*—for example in quartz (14), selenite and aluminium oxide, in the region of 7 μ to 10 μ (μ=0·001 mm.).

Modern conceptions of black body radiation are the result of slow development from such simple experiments as those of Draper[1] (4), who found that the interior of rifle barrel became luminous at a certain temperature; of Christiansen (6), who observed that the scratches and conical-shaped holes in an incandescent metal surface were brighter than the smooth surface; and of St. John (7), who found that the selective emission commonly observed in the rare oxides disappeared when these substances were heated to incandescence in a uniformly heated porcelain tube.

(i.) *Radiators.* — The modern uniformly heated cavity or so-called black body is the outcome of such observations and facts as those just mentioned. It is the invention of Wien and Lummer (8), and, for attaining high temperatures, it consists of a diaphragmed porcelain tube wound with a platinum ribbon which is electrically heated. In the modification used by Coblentz (15) this radiator consisted of three concentric porcelain tubes A, B, and C (*Fig.* 1). The inner tube A, of Marquardt porcelain, is wound uniformly with a platinum ribbon 0·02 mm. in thickness. This ribbon is 20 mm. wide in the centre, gently tapering to 10 mm. wide at the ends in order to provide extra heating at the ends so as to compensate for heat losses by radiation. The middle tube B is wound with a ribbon of platinum 10 mm. wide and 0·02 mm. in thickness. The ends are closely wound in order to provide extra heating and, in this manner, to compensate for the cooling of the ends by radiation. By properly regulating the current through these platinum strips the cavity can be heated uniformly throughout the greater part of its interior, as shown by Waidner and Burgess (16). The use of platinum ribbon instead of thick platinum wires eliminates local non-uniformity of temperature. The central 25 mm. length of the tube A was wound so that one turn of the platinum ribbon entirely covered what comprises the theoretical black body.

In radiators of this type the porcelain tubes sag on heating to 1200° C. To overcome this difficulty, in the determination of the radiation constants, to be described presently, Coblentz (15) placed a wedge-shaped fragment of porcelain with a sharp edge S (*Fig.* 1), about 2 mm. long, about 25 to 30 mm. to the rear of the radiating diaphragm. This support causes no local cooling and it preserves the life of the radiator. Further improvements are made in completeness of the emission of the radiation by painting the walls and the front side of the radiating diaphragm with a mixture of chromium oxide and cobalt oxide. Since these oxides become electrically conducting at 1200°, the front thermocouple, Th, was completely enclosed in a porcelain insulating tube, as shown in *Fig.* 2, B. The short piece of

[1] The number (4) and other similar numbers refer to the Bibliography at the end of the article.

porcelain which lies across the radiating wall of the black body was covered with cobalt oxide. Since the cobalt oxide did not adhere well to the porcelain tube, the latter was first covered with a thin layer of iron oxide, obtained by wetting the porcelain tube with writing ink which was burned into the tube. The cobalt oxide paint was then applied and also burned upon the tube by means of a blast lamp. After replacing the thermo-

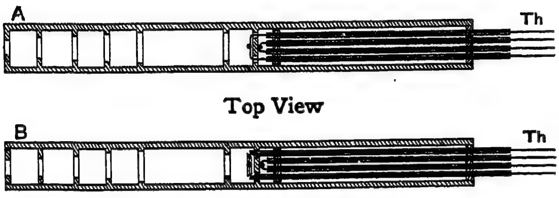


FIG. 2.

couple, enclosed in this short tube within the radiator, the whole interior was again painted with a mixture of cobalt and chromium oxides. In order to be able to insert the thermocouple including the short tube, and also to paint the interior of the radiator, the first diaphragm (which in the commercial porcelain tube has an opening of only 1 cm.) was removed (*Fig.* 2, B), and after making all the adjustments another porcelain diaphragm was put in its place.

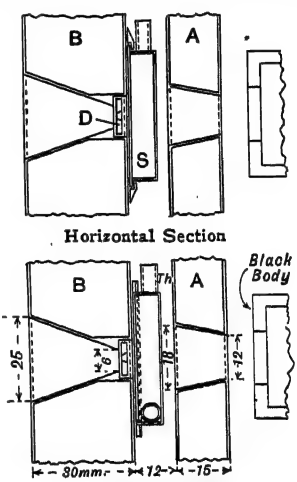
The cobalt oxide has a high temperature coefficient of absorption, so that it appears black on slight heating. Its emission spectrum (14) (*loc. cit.* p. 120) is continuous, so that there is less difficulty in producing a perfect radiator than when the radiating enclosure is of white porcelain (17, 18). However, porcelains having a low melting-point when heated above 1000° emit a far more continuous spectrum than the "Marquardt porcelain" from which are made the black bodies ordinarily used.

All these tubes, especially the painted ones, become electrically conducting when heated above 1300° C. Hence, at the highest temperatures the radiator should be heated by alternating current from a motor generator which is operated from a storage battery. By this means there is no more difficulty in maintaining a constant current than when heating the radiator directly by means of current from a storage battery. By means of a low-resistance rheostat at the observing table, the current can be regulated so that the temperature of the radiator can be kept to a few hundredths of a degree.

(ii.) *Water-cooled Shutters and Diaphragms.*—In discussing the relative merits of the various determinations of the radiation constants, and in particular the coefficient of total radiation, the writer will have occasion to refer frequently to the use, or misuse, of shields for preventing the receiver from being heated by the radiator, and of shutters for exposing the receiver to radiation. It is therefore important to consider briefly the functions of this apparatus.

The determination of the coefficient of total radiation usually consists in noting the heat interchange when the receiver is exposed to two radiators which are at widely differing temperatures, say 0° C. and 100° C., *i.e.* ice and boiling water. The radiator at the lower temperature may be and usually is used as a shutter. If the temperature of the shutter is lower than that of the receiver the latter radiates to the shutter. Hence it is important to have the receiver face a large (say water-cooled) diaphragm, maintained at a constant temperature, at the back of which is placed the shutter and the radiator. In this manner the surrounding conditions, facing the receiver, are not changed when the shutter is moved in order to expose the receiver to radiation.

There is no doubt that some of the unusual results obtained in the determination of the



Vertical Section

FIG. 3.

radiation constants are owing to the fact that the receiver did not face uniform conditions when the shutter was opened and closed in making the measurements.

Experience shows that the arrangement of the shutter should be similar to that illustrated in *Fig.* 3, which particular design of apparatus was used by Coblentz in the determination of the coefficient of total radiation described on a subsequent page.

The water-cooled shield employed consisted of a tank A, 25 cm. in diameter and 1·5 cm. in thickness, the radiator, and a tank B which faced the radiometer. The water-cooled shutter S (*Figs.* 3 and 6), consisting of a thin metal box 3·5 by 3·5 by 0·8 cm., was operated in vertical ways between these two shields. A mercurial thermometer was inserted at Th. Its bulb was in the water flowing through this shutter and was used to measure the temperature T used in equation (1). A more detailed illustration of a similar shutter is given elsewhere (15), and some of the parts are shown in *Fig.* 3. The side of the shutter facing the radiometer was smoked in a sperm candle, and in connection with the conical-shaped opening (painted black and smoked) in the water-cooled diaphragm B it formed a miniature black body the temperature of which remained constant throughout a series of measurements. The temperature of the shutter was easily kept constant to within 0·5° C. The opening in the diaphragm B, which must be accurately known, was defined by means of a series of small perforated discs of brass 8 mm. in diameter and 1·5 mm. in thickness, which were mounted in a recess provided for the purpose. These brass discs were provided with accurately cut, knife-edged holes which varied from 2 to 5·5 mm. in diameter. In this manner one could easily change the size of the opening which admits radiation upon the receiver and thus study the effect upon the radiation constant.

II. The Coefficient of Total Radiation

It has been shown by Stefan (1), Schneebeli (10), Wilson (11), Lummer and Pringsheim (9), Valentiner (12), and others (20), that the total radiation emitted from a uniformly heated cavity is proportional to the fourth power [1] of the absolute temperature. The measurements extend over a temperature range of about 1500° C. They verify the fourth power law within the accuracy of the temperature scale and the general errors of observation; these are small owing to the fact that the measurements are relative.

The evaluation of the coefficient of thermal radiation in absolute measure is an extremely difficult task which has been attempted by numerous experimenters whose data will now be considered.

The early determinations of the coefficient of emission of a substance (*e.g.* blackened plates, balls of glass or copper) by Lehnebach (19), Christiansen (21), and others are of great historical interest, but they cannot be considered in connection with present-day determinations of the coefficient of total radiation of a uniformly heated enclosure.

[1] See "Pyrometry, Total Radiation," Vol. I.

§ (3) Bolometric Methods of Electric Compensation. (i.) *Observations of Kurlbaum* (22).—The first determination of the constant of total radiation in absolute value, by a method which is free from the gross experimental errors that are inherent in the work, was made by Kurlbaum.

The principle of the method is as follows: Three branches of a Wheatstone bridge (bolometer) consist of thick manganin wires, which are not affected by a change in current through the bridge. The fourth branch consists of thin strips of platinum, the resistance of which is affected by a change in current in the bridge.

Starting with the bridge balanced, the bolometer branch of thin platinum, at a distance of 18 to 20 cm., was exposed to the radiation from a black body at 100° C., and the change in resistance (or galvanometer deflection) was noted. With the bolometer branch shielded from radiation, the bridge current was varied until the change in resistance in the bolometer branch was equal to that attained when exposed to radiation. In other words, the bridge current was increased by a sufficient amount to produce the same galvanometer deflection as was caused by radiation falling on the bolometer strip. From a knowledge of the change in the bridge current, the bolometer resistance, etc., Kurlbaum was able to compute the energy input. He used two very different surface bolometers and found very concordant values. He found a value of $\sigma = 5{\cdot}32$. In a subsequent communication (23) he made a correction of 2·5 per cent for loss by reflection from the platinum black surfaces of the bolometer. This gives a value of $\sigma = 5{\cdot}45$. Coblentz (24) determined the reflecting power of platinum black, and found that, using the blackest obtainable deposits, the reflecting power is almost 2 per cent for wave-lengths at 8 to 9 μ. An examination of six surface bolometers (12 branches or 24 surfaces of about 3 by 4 cm. area), purchased abroad, showed that all of them had microscopically small bright patches of bare platinum. These bright areas would increase the loss by reflection. From this it appears that Kurlbaum's correction to his work is none too large.

(ii.) *Observations of Valentiner* (12).—The method of measuring the coefficient of total radiation by means of a large surface bolometer was continued by Valentiner, who found quite different values, depending upon the blackness (kind) of radiators employed, which were heated as high as 1450° C. For the two blackest radiators (the steam-heated radiator "W.S.K." and the electrically heated unblackened porcelain tube radiator "G.S.K.") he found a value of $\sigma = 5{\cdot}36$.

In a subsequent paper (25) he made a correc-

tion of about 4 per cent for various causes (*e.g.* 1 per cent for lack of blackness of the radiator, 2·5 per cent for reflection from the bolometer, etc.) giving a value of $\sigma=5{\cdot}58$. No correction was made for atmospheric absorption.

The numerical values of the various determinations made by Valentiner, upon any one radiator, vary by about 3 per cent, which is greater than the effect of atmospheric absorption in a space of 30 to 50 cm. However, when using a larger porcelain tube radiator (12), "G.S.K.," the distances were longer than usual (being respectively 89 and 125 cm.) and the effect of absorption seems unmistakable. The value of the radiation constant is about 2 per cent smaller for the greater distance, *i.e.* an increase of 35 cm. in optical path introduced an absorption of about 2 per cent.

A conservative estimate of the correction to Valentiner's data for absorption by atmospheric water vapour and carbon dioxide is 2 to 3 per cent. This increases Valentiner's value to $\sigma=5{\cdot}68$ with a possibility of the value being as high as $\sigma=5{\cdot}75$, which is in remarkably close agreement with the various determinations made by other methods. That a correction of not less than 1 per cent is to be applied to these data is indicated in Gerlach's (33) recent work on atmospheric absorption of black body radiation.

The surface bolometers used by Kurlbaum and by Valentiner consisted of two grids of thin platinum foil, placed one back of the other, with the edges overlapping. In this manner a continuous surface is formed, but part of the rear bolometer strips are shielded from the radiation incident on the front strips. The operation of the receiver is unsymmetrical. Gerlach (28) has discussed the effect of this overlapping, and has shown that in some cases it might lower the value of the observed radiation constant by a very considerable amount. Coblentz (27) has found that using one bolometer strip back of the other, 50 per cent of the energy, incident upon the first receiver, is re-radiated to the rear one. In the reverse process, of heating the strip electrically, this effect would enter symmetrically and a low value would result.

It was found by Kurlbaum and by Valentiner that the rate of temperature rise was different when heating the bolometer by absorption of radiation and by electrical heating. The low values of the radiation constant obtained by Valentiner are explained (28) in part as the result of reading the galvanometer deflection before temperature equilibrium was fully attained.

Callendar (30) has described surface bolometers which would overcome some of the defects of conduction, etc. However, using radiators at 600° to 1000° C., there is no need to employ such large receivers with the unavoidable defects just noted.

(iii.) *Observations of Gerlach* (31).—The following method is not exactly bolometric in principle, but it may be considered at this point since its virtues have been under discussion in connection with the results obtained by the bolometric methods just mentioned. Gerlach determined the constant of total radiation, σ, by a modification of Ångström's electric compensation pyrheliometer.[1] In the original pyrheliometer of Ångström (32) the receivers are two thin narrow sheets of manganin to each of which is attached one junction of a thermo-element which is joined through a galvanometer. One of these manganin strips is exposed to solar radiation. Through the other manganin strip is passed an electric current of such strength that it is heated to the same temperature as is the receiver which is exposed to solar radiation. The equality of temperature is indicated when there is no current flowing through the galvanometer.

In the instrument as used by Gerlach (31) the receiver consists of but one strip of manganin. At the back of this manganin strip and close to it is a thermopile of 45 elements (joined through a galvanometer) which is heated by radiation from the strip, thus differing from Ångström's device, in which a single thermo-junction is heated by conduction from the manganin strip. The object of using many thermo-junctions is to integrate the radiation from the whole strip, thus eliminating any irregular heating caused by inequalities in the thickness of the receiver, which was blackened electrolytically with platinum black. The manganin strip is heated electrically to the same temperature as that attained by the strip when exposed to the radiation from a black body, which was heated by steam. From a knowledge of the resistance of the strip and the electric current the energy input, and hence the value of the radiation constant, is determined.

Gerlach experienced some difficulty in determining when he had exact compensation when heating the receiver electrically and radiometrically. Covering the sides of the receiver with knife-edged slits had no effect upon the radiation constant; but shielding the ends of the manganin strip from radiation caused the value to increase from $\sigma=5{\cdot}83$ for the unshielded ends to a constant value of $\sigma=6{\cdot}14$, when the shields covered 1·5 mm. or more of the ends of the strips. This is caused by heat conduction from the receiver to the heavy copper electrodes. In practice he exposed the whole length of the receiver to radiation, claiming that the heat conducted from the ends is the

[1] For an account of this and some others of the instruments referred to see Vol. III., "Radiant Heat, Instruments for Measurement of."

same when the metal strip is heated radiometrically and electrically. This appears to be the weak point in the device. In the instrument used by Coblentz (20, 29) this difficulty was eliminated by placing the potential terminals on the receiver, and at some distance from the heavy electrodes. This seems to be the logical way to use such a receiver.

The work of Gerlach aroused considerable discussion among experimenters previously engaged on this same problem. This brought forth a very considerable amount of laborious experimental work by Gerlach (35), in which he showed that his apparatus gave the same values when operated by Kurlbaum's (22) bolometric method as when operated by his method.

In subsequent investigations Gerlach determined the effect of atmospheric absorption (33) and of blackening the receivers (36) upon the radiation constant. He operated the radiator at higher temperatures (to 450° C.), but the radiometric apparatus remained the same as in the earlier work. With this outfit his new value (37) of the coefficient of radiation, uncorrected for reflection of radiation from the receiver and for atmospheric absorption, is decidedly lower than that which he had previously obtained under similar conditions, being of the order $\sigma = 5{\cdot}6$ to $5{\cdot}7$. Applying a correction of about 1·7 per cent for reflection of platinum black as observed by Coblentz (24) and depending upon the temperature of the radiator, a correction of 0·2 to 1·2 per cent for absorption of CO_2, the mean value of 52 independent sets of measurements is $\sigma = 5{\cdot}85$. In a recent discussion of his data Gerlach (102) places his value at $\sigma = 5{\cdot}80$.

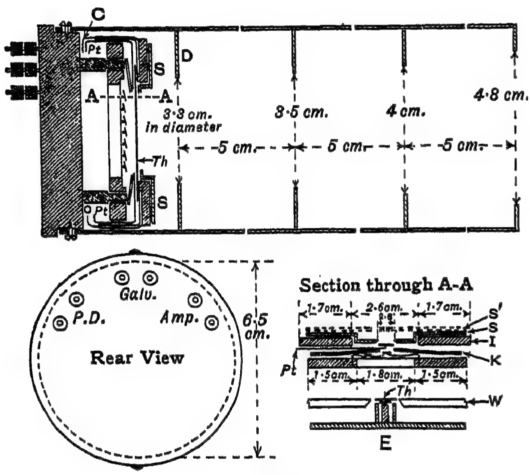


FIG. 4.

Gerlach's determination is one of several, which should be a convincing demonstration that the extreme values of $\sigma = 5{\cdot}3$ to $6{\cdot}5$ are subject to correction, in so far as the corrections can be determined.

(iv.) *Observations of Coblentz* (20, 29, 39).—This research was undertaken after a study of the instruments and methods used in the earlier determinations, and an effort was made to embody the good features and eliminate the defective ones previously employed. It is therefore proper to discuss the work in considerable detail.

In designing the apparatus an attempt was made to embody black body conditions in the radiator, *Fig.* 1 (including the shutter, *Fig.* 3), and the receiver.

The lack of blackness of the radiator has been discussed by Wien and Lummer (8), who give a method for computing the correction for the opening in the radiator, on the assumption that the enclosure is spherical and diffusely reflecting. The amount of energy that can escape by diffuse reflection through the opening is determined practically by the size of the opening as compared with the total area of the interior of the radiator. In the porcelain tube radiator used here the interior is uniformly heated over a length of 8 to 10 cm. However, for the purpose of the present computation, a length of only 2·5 cm., which is defined by the first diaphragm, is considered. The area of the enclosure is 37·6 cm.^{2} and that of the opening is 3·1 cm.^{2}. On the basis that the reflecting power of the interior of the painted radiator is 7 per cent (*loc. cit.* (20), p. 523, Table 3) the loss of energy by diffuse reflection through the inner diaphragm is 0·63 per cent. Using an unpainted Marquardt porcelain radiator the coefficient of total radiation is decreased by about 1 per cent (*loc. cit.* (29) p. 571), showing that the question of lack of blackness of the radiator is important.

For a receiver (40) it was decided to use a modified Ångström pyrheliometer (32) embodying novel features which had not yet been tried by others. In order to reduce the heat capacity of the manganin receiver Th, *Fig.* 4, and provide better insulation, the thermopile of bismuth and silver, having a continuous receiving surface, was placed a short distance back of the receiver as shown in E, *Fig.* 4. The apparatus used by Gerlach embodied this same idea.

The crucial and novel part of the apparatus was a receiver R, *Fig.* 5, with potential ter-

minals PP attached thereto, at a sufficient distance from the ends to avoid the question of heat conduction to the electrodes. These potential wires, which were from 0·003 to 0·025 mm. in diameter, accurately defined the length of the central part of the receiver which was utilised in the measurements. By

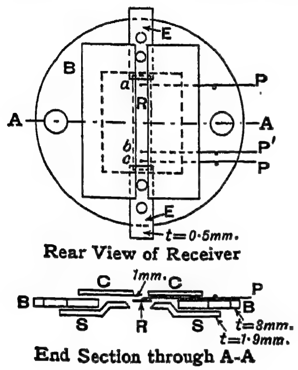


Fig. 5.

exposing the whole length of the receiver to radiation, conduction losses from its end do not enter the problem.

The effect of the presence of the potential terminals was determined (20) (by using a third terminal P′) and found negligible, viz. 0·3 per cent. The method of finding the difference of potential between two terminals attached to the receiver seems more certain than to find the difference of potential between two heavy electrodes as used by Gerlach (31).

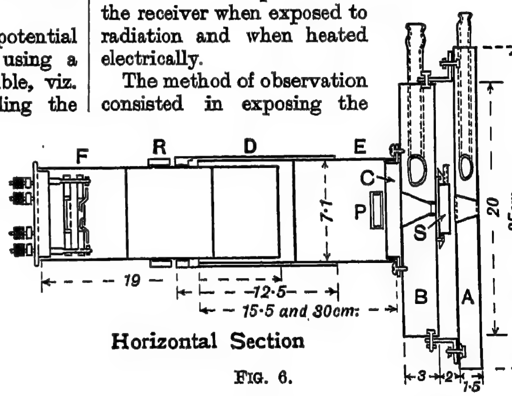
The operation of this type of radiometer, with a hemispherical mirror placed in front of the receiver (20, 24, 38), may introduce errors when the receiver is heated by electrical means. Hence, after conferring with specialists in geometrical optics who were in agreement with the writer's opinion that a reflecting enclosure is likely to introduce errors, the receiver was used without the hemispherical mirror. Subsequently this mirror was employed in a separate experiment to determine the diffuse reflection (24) from lampblack and platinum black, and finally in determining the loss by reflection from some of the receivers actually used in the observations. Hence, the loss of radiation incident upon the receivers was probably accounted for as accurately as it would have been by employing a "blackening" device in front of the receiver.

In order to test the question of the accuracy of the corrections used for eliminating the loss by reflection, a series of observations was made on one receiver. In this test the slits in front of the receiver, and all other conditions remained unchanged. The only variation was in smoking the platinum black receiver with a sperm candle after making the first set of observations. The reflecting power of platinum black is taken to be 1·7 per cent (39) and that of lamp-black 1·2 per cent. The respective determinations, after correcting for energy lost by reflection, were $\sigma = 5{\cdot}822$ and $\sigma = 5{\cdot}814$. They differ by only about 0·1 per cent, which is very satisfactory, and shows that the reflection factors were well determined.

The assembled apparatus is shown in *Fig.* 6, in which A and B are the water-cooled diaphragms, S the shutter (see *Fig.* 3), F the radiometer, and D the telescoping diaphragmed tubes enclosing the optical path from which the moisture was removed by means of phosphorus pentoxide P. Subsequent tests showed that in this outfit the effect of atmospheric absorption was less than 0·1 per cent (20), which conclusion is substantiated by recent measurements made by Gerlach (33).

The thermopile was connected with an unusually well shielded iron-clad Thomson galvanometer of special design (38, 41), which served merely as a null instrument to indicate the rise in temperature of the receiver when exposed to radiation and when heated electrically.

The method of observation consisted in exposing the receiver to radiation and noting the galvanometer deflection; then in heating the strip electrically to give, within 1 per cent, the same deflection. The measurements of the electric power put into the strip were made with the same potentiometer used in measuring the temperature of the radiator. No difficulty was experienced in determining temperature equilibrium on heating the receiver by radiation and by electrical means.



Horizontal Section

Fig. 6.

The exact amount of energy necessary to cause the same deflection as that produced by absorbing radiant energy is obtained by multiplying the observed energy input by the ratio of the galvanometer deflections. This gives the "constant" of each receiver. In order to reduce all the measurements to a common

basis, and at the same time obtain a value of the coefficient, or the so-called Stefan-Bolzmann constant, of total radiation, the custom of previous experimenters was followed in reducing the data. For this purpose it is necessary to know the area A_1 of the water-cooled diaphragm, the area A_2 of the receiver which is exposed to the radiation emanating from A_1, the distance D between these two surfaces, the absolute temperature T_1 of the shutter, and the absolute temperature T_2 of the radiator. The electrical energy input necessary to produce the same (galvanometer deflection) rise in temperature as was produced by exposing the receiver to the radiator is EI, both of which quantities are measured with the potentiometer as already described. Under these conditions the energy consumed by the electrical stimulus may be equated to the energy emanating from the radiator, or

$$EI=\frac{\sigma}{\pi}(T_2^4-T_1^4)\frac{A_1A_2}{D_2}, \quad . \quad . \quad (3)$$

where σ is the coefficient of total radiation, E is the voltage, and I is the current.

Usually the distance between the two surfaces A_1 and A_2 is so close that experimenters have had to apply a second term correction (42). In Gerlach's (31) work the correction applied was

$$\frac{A_1A_2}{D^2}\left(1-\frac{1}{6}\frac{a^2+b^2+a_1^2+b_1^2}{D^2}\right), \quad . \quad . \quad (4)$$

where the sides of the openings in the rectangular diaphragms were a, b, a_1, b_1. A shorter correction factor is

$$\frac{A_1A_2}{D^2}\left(1-\frac{1}{4}\frac{a^2+b^2}{D^2}\right), \quad . \quad . \quad . \quad (5)$$

which results from neglecting the effect of the first diaphragm, which was very small in this work. In either case this second term correction is quite negligible in the present investigation, since it usually amounts to less than 2 parts in 1000.

The radiator was operated at a temperature of 800° to 1100° C., within which range the temperature scale is quite well determined.

To the data obtained with 10 receivers, a correction of 1·2 per cent, for losses by reflection, was applied to measurements made with receivers covered with lamp-black (soot) and a correction of 1·7 per cent to measurements made with receivers covered with platinum black. The reflection from a receiver covered with platinum black, then smoked, is 1·2 per cent. As just mentioned, these corrections were determined by direct measurements upon some of the receivers used in this investigation and by comparison of the surfaces of the other receivers with samples of lamp-black whose reflection losses had been determined in a previous investigation (24).

The value obtained for the coefficient of total radiation, after applying the corrections just mentioned, is

$$\sigma=(5{\cdot}722\pm0{\cdot}012)\times10^{-12}\text{ watt cm.}^{-2}\text{ deg.}^{-4}.$$

No correction was made for absorption by atmospheric carbon dioxide, which was found to be less than 0·1 per cent. According to Gerlach's (33) measurements this might perhaps amount to 0·1 or 0·2 per cent, thus increasing the value to $\sigma=5{\cdot}73$.

The method of operation is unsymmetrical in that, when the receiver is exposed to radiation, the heating is produced in the lamp-black surface, while in passing an electric current through the strip the heat is generated within the receiver. However, from the data obtained with receivers differing 10 times in thickness, and covered with different kinds and thicknesses of absorbing material, it appears that the manner of heating the receiver has but little effect upon the final result.

For any one receiver, operated under different conditions, the precision attained is usually much better than 1 per cent. For the different receivers the maximum range in the value of σ is of the order of 1·5 to 2 per cent. This seems to be independent of the length and width of the receiver, and of the kind of slits used. The accuracy attained with this method of evaluating energy in absolute measure, as estimated by the departure of individual determinations from the mean value, appears to be of the order of 1 per cent. To this extent one can consider the present device a primary instrument for evaluating radiant energy in absolute measure.

(v) *Observations of Kahanowicz* (44).—The apparatus used by Kahanowicz is essentially a modification of the Ångström pyrheliometer (32).

The receiver was placed at the centre of a spherical mirror with an opening in one side to admit radiation. In this manner the correction for reflection was eliminated. The shutter was close to the receiver. If its temperature was different from that of the water-cooled diaphragm, which was before the radiator, errors in the radiation measurements would occur. As mentioned elsewhere in this paper, the shutter should be placed between the water-cooled diaphragm and the radiator, to avoid a change in the surroundings facing the receiver when the shutter is raised for making the radiation measurements.

The temperature range was from 260° to 530° C. The distance from the radiator to the receiver was 35 to 55 cm. A series of 28 measurements gave an average value of $\sigma=5{\cdot}61$. Of this number 11 gave a value of $\sigma=5{\cdot}7$. Out of a series of 4 measurements made in December 1916 (lower humidity), with the distance $\sigma=56$ cm., 3 gave a value of $\sigma=5{\cdot}7$.

A recent determination, using temperatures ranging from 256° to 1076° C., gave a value of $\sigma=5\cdot61$.

No corrections were made for atmospheric absorption, which for the temperatures used is not negligible. In a previous paper (*loc. cit.* (20), p. 576) it was shown that on removing the moisture (vapour pressure of 10 to 12 mm.) from a column of air 52 cm. in length, the radiation constant was increased from $\sigma=5\cdot41$ to $5\cdot55$, or about 2·6 per cent. The vapour pressure at Naples is considerably higher than at Washington. From these (39), as well as Gerlach's (33) data, it appears that the corrections for atmospheric absorption should be at least 1 per cent. For the temperatures at which the radiator was operated, a conservative estimate of the correction to the radiation data obtained by Kahanowicz is 1·5 to 2 per cent, resulting in a value of $\sigma=5\cdot69$ to 5·72. In other words, the Naples value of the coefficient of total radiation is comparable with other recent determinations which indicate a value of $\sigma=5\cdot7$.

(vi.) *Observations of Puccianti* (43).—The method of operation is just the opposite of that usually employed, and in view of the small radiant energy exchanges involved in comparison with the transfer of energy by heat conduction and by convection, this may perhaps be the weak point in this method.

Puccianti constructed a bolometer in the form of a black body, which is kept at room temperature. This black body is really the radiator. The other black body (which is really the receiver), instead of being at a higher temperature, as usually is the case, is at the temperature of carbon dioxide, snow, or of liquid air. He measured the electric power which had to be supplied to the first black body to keep the temperature constant in order to compensate for the energy lost by radiation to the second black body. He very ingeniously constructed two black-body bolometer branches exactly alike, the one to be exposed to the cold receiver and the other to be protected from it. Each of these bolometer branches consisted of a vessel of 0·1 mm. sheet copper having the form of a cone and a frustum of a cone united at the bases as shown in B_1 and B_2, *Fig.* 7. The lengths were 12 cm.; the maximum internal diameter was 4 cm. The internal surface was smoked. The external surface was polished, and upon it was wound two thin insulated wires. One of these wires, of iron, formed the bolometric branch, and the other wire, of manganin, was used as a heating resistance. The other two branches of the bolometer bridge were formed of resistance coils and the whole was connected with a galvanometer and storage battery as in any ordinary bolometer. The two sensitive black-body branches of the bolometer were in an evacuated vessel C_1 which was kept in a tank of water.

The receiver was a blackened glass bulb N_1 immersed in liquid air, and the bolometer was allowed to radiate to this receiver. The constant K of the instrument was determined from the diameters of the diaphragms D_1 and D_2. A correction was made for the energy interchange between these two openings (42).

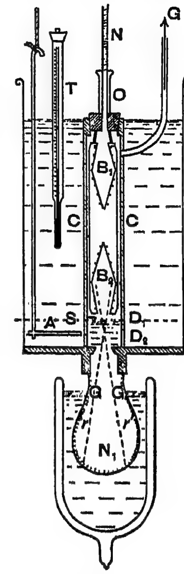


FIG. 7.

Puccianti assumed that the shutter and the bolometer branch B_2 were at the temperature T of the water bath. The resistance R of the manganin heating coil surrounding the bolometer branch B_2 was determined. In the course of the test Puccianti measured the voltage E, which was necessary for compensation to prevent the bridge from being unbalanced when the branch B_2 was exposed to the receiver N_1 at the temperature T_0. He obtained a value of $\sigma=5\cdot96$.

The method is an ingenious variation from the usual procedure. The apparatus should have been constructed so that both bolometer branches could have been used as radiators. From the illustrations it appears that radiation from one branch could fall upon the other, which would introduce errors. Another uncertainty is the temperature and the manner of operation of the shutter. Tests might have been made to determine whether the bolometer remained balanced when a heating current was applied to both branches, without allowing one branch to radiate to the receiver. Furthermore, to repeat the herein oft-mentioned device, a heating coil should have been inserted temporarily within the radiator to determine the energy input as compared with the energy input required in the outer heating coils in order to maintain a balance. The device, as used, is unsymmetrical in that the heating coil is not in the proper place to operate most efficiently. From this it appears that the constant should be smaller than that indicated by his measurements.

§ (4) THERMOMETRIC METHODS WITH "BLACK" RECEIVERS. (i.) *Observations of Féry* (45).—In order to eliminate the question of reflection from the surface Féry (45) made

a series of determinations of the radiant constant σ, by means of a thermojunction, formed into a long conical-shaped metal receiver which was blackened on the inside. On the outside of the cone, and insulated from it, was wound a heating coil of known resistance for calibrating the receiver. This was done by noting the temperature rise (galvanometer deflections) with the energy input, in watts. The receiver was then exposed to an electric furnace which was heated to various temperatures from 500° to 1200° C., and the corresponding galvanometer deflections were observed. He obtained a value of $\sigma=6{\cdot}30$.

(ii.) *Observations of Féry and Drecq* (46).—The work was undertaken anew by Féry and Drecq (46), the receiver being a cone of brass, having an aperture of 30°, placed within a large sphere of brass which was surmounted by a glass tube with a capillary 1 mm. in diameter. The large brass sphere was filled with alcohol, the whole forming a thermometer in which 1 mm. rise in height of the column in the capillary indicated a rise in temperature of 0·005°. Surrounding the outside of the brass case was a coil of wire through which an electric current was passed and the energy input noted which caused the same temperature as that produced by exposing the opening of the cone to the radiator. They found a mean value of $\sigma=6{\cdot}51$, which is the highest value yet obtained. The relative values obtained with the instrument show a fair constancy over the whole temperature range, which extended to the melting-point of gold. However, the same is true of Valentiner's values determined by the bolometric method. It indicates that there is some constant factor which is not eliminated and which causes a high or low value.

Various suggestions have been made as to the cause of their excessively large values. It appears that part of the difficulty lies in the unsymmetrical way in which the instrument is operated. It is calibrated by a heating coil which is in connection with the alcohol (or air, in the first case) and can warm the latter by conduction and by absorption of radiant energy. On the other hand, the incoming radiation must be transformed by absorption in the cone and then reaches the alcohol principally by conduction. As a result of this type of calibration, the apparent value of the radiation constant is higher than the true value. The weak point lies in not having the heating coil within the receiver, which should be so constructed that but little, if any, of the entering radiation, or the energy radiated from the heating coil, can escape through the opening in the receiver. This might easily have been done by forming the receiver into a double-coned receiver, such as was used by Puccianti. However, even with all these precautions, this type of receiver does not appear satisfactory.

Another determination of this constant was made by Féry and Drecq (47) in which the radiations from an electric furnace fell upon a strip of platinum (area 36 by 55 mm. and 0·03 mm. in thickness), blackened electrolytically with platinum black. The radiation measurements were made by sighting upon the front and the rear surfaces of the platinum strip by means of a Féry pyrometer, the angle of incidence being 48°. Their new value of the radiation constant was $\sigma=6{\cdot}2$. The measurements upon the posterior surface resulted in a value of $\sigma=5{\cdot}57$, which is said to correspond with the measurements on the anterior surface made with plane receivers. If we correct the latter value by (the round number) 2 per cent for reflection (24), we obtain the value $\sigma=5{\cdot}68$, which is of the same magnitude as observed by the writer. This new determination of σ by Féry and Drecq appears to have been made defective by their reduction of the original observations; for example, they claim that the coefficient of absorption of the receiver was only 0·82 to 0·84, which seems impossible from numerous and diverse experiments on the diffuse reflecting power of platinum black made by others. Attention has been called to this fact by Bauer (48), who placed their value of σ between 5·1 and 5·8, and, by making a correction of 2 per cent for reflection, deduced a value of $\sigma=5{\cdot}68$.

In all of the foregoing methods the data are meagre as to elimination of the various errors which may occur. For example, a source of error may arise in determining the power put into the platinum strip used in the last method. Another important source of error lies in the manner of operation of the water-cooled shutter, which should be placed between the water-cooled diaphragm and the radiator.

(iii.) *Observations of Bauer and Moulin* (49).—In order to obviate the difficulties encountered by Féry and Drecq (47) in calibrating their conical-shaped receivers, Bauer and Moulin calibrated their receiver (which was a Féry pyrometer) by sighting it upon a strip of platinum, heated to different temperatures by an electric current. In order to determine the amount of radiant energy falling upon the receiver it was necessary to eliminate the losses from the strip by conduction and convection. For this purpose the platinum strip was heated in air and in a vacuum, the power consumed being determined for a definite length of platinum, defined by potential terminals welded to the strip. Having calibrated the pyrometer by noting the galvanometer deflections for the various amounts of energy (in watts) put into the platinum strip, they sighted the pyrometer upon a black body heated to various temperatures and noted the

galvanometer deflections. Their first announced value was $\sigma=6\cdot0$. However, as in numerous other cases herein cited, errors were introduced in the final reduction of their data. They had observed the radiation emitted from the platinum strip at an angle of 13° from the normal and applied a correction (50) of about 12 per cent, which reduced their value to $\sigma=5\cdot3$. This correction is recognised to be much too large (48), so that their value of the radiation constant lies between $\sigma=5\cdot3$ and $6\cdot0$. They made no correction for atmospheric absorption, which would increase their value from $\sigma=5\cdot3$ to $\sigma=5\cdot6$ or $5\cdot7$. In an earlier communication on the solar constant, Bauer and Moulin (*loc. cit.* (50), p. 1658), using an Ångström pyrheliometer, found the value $\sigma=5\cdot7$.

(iv.) *Observations of Puccianti* (51).—In a continuation of his investigations, Puccianti gives a determination in absolute measure of the radiation of a black body in which the temperature change is measured by means of a toluene thermometer, the bulb of which is formed into a hollow cone that is allowed to radiate to a black-body receiver, which is at the temperature of liquid air. The measurement is made by compensating the heat lost by the thermometer by the application of an electric current.

The apparatus not being differential in construction, the temperature of the water bath had to be kept rigorously constant in order to have the meniscus of the thermometer move slowly and regularly. The response of the apparatus was, of course, slow and sluggish, which is a common property of this type of receiver (radiator), so that it required from four to eight minutes to obtain a measurement.

The same criticisms apply to this instrument that have been mentioned in the cruder form of thermometer used by Féry and Drecq (46). The energy for compensation is supplied by a heating coil which is in contact with the liquid (a good scheme in so far as it applies to heating the liquid) and on the side of the wall of the receiver opposite to that upon which the incoming radiations impinge. The arrangement is therefore unsymmetrical. The heat of compensation should have been supplied by a coil inserted within the receiver, provision being made that little or none could escape by reflection and by direct radiation through the opening. In this manner the condition of symmetrical heat interchanges would have been the most closely fulfilled. In the instrument as used the opportunity for escape of energy seems greater, so that in compensation there is a tendency to produce a value which is higher than the true value for radiation unaccompanied by conduction. By placing the heating coil within the receiver and using a high temperature radiator, Puccianti's device should be applied in the manner recently used by Keene (52).

Puccianti considered the precision of this method as high as that of the bolometric apparatus, but the sensitivity of the thermometric apparatus was very much inferior to that of the bolometer. Nevertheless, he seems to prefer the thermometric method in spite of its small sensitivity. He assigned a value of $6\cdot00<\sigma<6\cdot3$, and his intermediate value is $\sigma=6\cdot15$.

(v.) *Observations of Keene* (52).—The most recent determination of the constant σ is by Keene. The radiator consisted of an electric furnace which could be heated to 1000° C. The receiver consisted of a hollow spherical double-walled thermometer bulb provided with a small aperture in its side to admit the radiation to be measured, as shown in *Fig.* 8. The space between the walls is filled with aniline, which served as thermometric substance, its expansion being observed in a capillary tube in the usual way, 1 mm. division $=0\cdot0005°$ C.

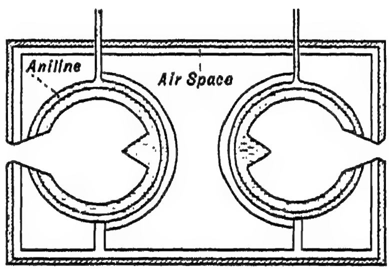


FIG. 8.

In order to eliminate the effect of the variation of room temperature, two such thermometers were used differentially, radiation being admitted into one of them, the differential effect giving a measure of the energy supplied. The interior of the bulb receiving the radiation was provided with an electric heating coil for the purpose of calibration. His paper contains the derivation of an exact expression for the energy interchange between two radiating coaxial circular openings; for he found that the approximate formula which is ordinarily used when the distance between the openings is large was not accurate enough for the work.

His value of the radiation constant is $\sigma=5\cdot89$.

The receiver and the radiator being close together, and the time for attaining temperature equilibrium being rather long, there is a possibility of diffusion of hot gases into the receiver when the shutter is raised for observa-

tions. This would tend to give a high value. It would have been interesting to determine if it would have required less electric power in the heating coil, provided the opening in the receiver had been closed with a reflecting cover to prevent escape of heated gases, while calibrating the device.

§ (5) INDIRECT AND SUBSTITUTION METHODS. (i.) *Observations of Shakespear* (53).—His value of the constant σ is obtained by a method which is based upon the principle that a heated body in air, in surroundings at a lower temperature, loses heat (*a*) by conduction, (*b*) by convection, and (*c*) by radiation. If the rate of loss by a body be observed in two cases, the only difference being that the emissivity of the radiating surface varies, other conditions remaining the same, it is quite correct to assume that the losses (*a*) and (*b*) will be the same and that the difference between the observed rate of loss of energy in the two cases is due only to the difference in the losses by radiation. If, now, these two different surfaces, at the temperature of boiling water, be exposed in turn to a radiomicrometer at room temperature, we obtain the ratio of the rates of the energy radiated by the two surfaces.

In the experiments, a plate of metal with a silvered surface was heated electrically to 100° C. and close to it was another plate, blackened with soot, which was cooled with water. Between the plates was air at atmospheric pressure. Shakespear measured the energy input in order to maintain the plate at 100° when (1) the surface of the plate was highly polished, and (2) when it was blackened. He measured also the ratio of the emissivities of the plate under these two conditions, using a radiomicrometer for the purpose. From this he obtained a value σ' in absolute units. He then compared the emissivity of the lamp-black surface at 100° with that of a black body at the same temperature by means of a radiomicrometer. From this latter comparison combined with the value of σ' he found a value of $\sigma = 5{\cdot}67$.

From this description it may be noticed that the essential parts of the method differ from that of Westphal (55) in that the radiator was a flat metal plate used in air instead of a vacuum, and that the black body, with which the emissivity of the plate had to be measured, was separate from it; while in Westphal's instrument the black body was self-contained within the metal (in the form of a cylinder) of which the emissivity of the surface had to be measured.

(ii. *Observations of Todd* (54).—In his experiments on the thermal conductivity of gases, Todd used a thin layer of air enclosed between two horizontal, parallel, good-conducting plates, which were maintained at different temperatures. The colder plate, of course, receives heat by radiation and by conduction through the air from the hotter plate, which is above it. Communication from the surrounding air is shut off by an insulating ring, and the two plates being close together, in comparison with their linear dimensions, the convection currents are eliminated. He determined the energy lost by radiation by varying the distance x between the two plates and noting the corresponding variation in the quantity Q of heat passing from the upper to the lower plate. These values of x and Q when plotted from a rectangular hyperbola and the horizontal asymptote give the value R of the radiation. The energy input was determined by a calorimetric method. In order to determine the constant σ he had simply to find the ratio of emissivities of the blackened plate to that of a black body at the same temperature, for which purpose a radiomicrometer was employed. The value of this ratio and the constants obtained in the main part of his experiment enabled him to compute the radiation constant, which he found to be $\sigma = 5{\cdot}48$.

This, like the preceding method, is likely to give a low value of the coefficient of radiation owing to uncertainty of the exact temperature of the radiating surface of lamp-black. A thin layer of soot is fairly efficient in its emission and absorption. But a thick layer of soot is quite non-conducting of heat, as was found in the measurements of diffuse reflection (34).

(iii.) *Observations by Westphal* (55).—An important determination of the coefficient of total radiation was made by Westphal (55). The experiment consisted in comparing the emissivity of a cylindrical block of copper, when it was highly polished and when it was blackened, with the emissivity of a black body at the same temperature. The novelty involved in the method is in having the black body contained within the cylinder as shown in *Fig.* 9. The copper cylinder was heated electrically, and to reduce the energy losses by gaseous heat conduction this copper cylinder was suspended in a glass flask from which the air could be exhausted to 1 mm. pressure. The outer surface of the cylinder was either highly polished to give it a low emissivity E_2 or painted with lamp-black to give it a high emissivity E_1. The end surfaces remained unchanged. The heat losses by conduction and convection were therefore the same throughout the experiment, and the difference in energy input, when the surface of the cylinder had a high emissivity and when it had a low emissivity, was a measure of the energy lost by radiation.

If the temperature of the glass flask is kept constant T_0, and the blackened cylinder is heated

electrically to the (absolute) temperature T, then the energy input W which is required in order to maintain a stationary temperature is

$$W_1 = O\sigma E_1(T^4 - T_0^4) + f(T,T_0). \quad . \quad . \quad (6)$$

In this equation, O is the surface of the cylinder and (T,T_0) is an unknown function which represents the

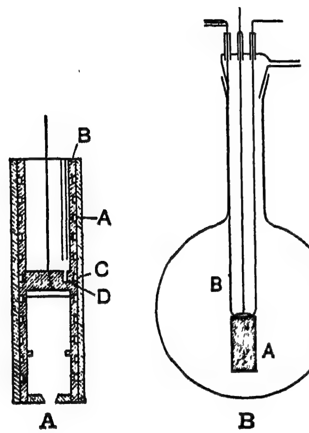


FIG. 9.

energy lost by conduction through the air and lead wires, and by radiation from the ends of the cylinder.

If now the surface of the cylinder is polished and thus given the emissivity E_2, and if the cylinder is heated to the temperature T, while other conditions remain unchanged, the energy lost from the cylinder is

$$W_2 = O\sigma E_2(T^4 - T_0^4) + f(T,T_0) \quad . \quad . \quad (7)$$

and the constant sought is

$$\sigma = \frac{W_1 - W_2}{O(E_1 - E_2)(T^4 - T_0^4)}. \quad . \quad . \quad (8)$$

Using the surface of the cylindrical body when highly polished, Westphal proceeded to find the curve of watt-input W_2 of the body between the temperatures 350° and 425° absolute as a function of the temperature. Then the surface of the body was brought to a high emissivity by applying in succession different blackening materials, and the energy W_1 input was measured at different temperatures.

The emissivities of the surface and of the interior of the copper cylinder were compared by means of a thermopile. The numerous details need not be discussed. Suffice it to say that the work appears to have been thoroughly done, and from the nature of the method it seems free from gross systematic errors. His mean value is $\sigma = 5{\cdot}54$.

He modified the apparatus and extended the observation over a wider range of temperature with a view to increasing the accuracy. The new value (56) agrees well with the earlier determination, being $\sigma = 5{\cdot}57$. Although nothing further seems to have been published subsequently, Westphal obtained further measurements which yielded the value $\sigma = 5{\cdot}67$ (58). This is in excellent agreement with Shakespear's results (53) showing that the coefficient of radiation as determined by reliable methods is of the order $\sigma = 5{\cdot}7$.

As already noted in discussing Todd's data, it is uncertain whether in all cases the surface of a layer of black soot has the temperature of the metal plate upon which it is deposited. Gerlach (36) performed experiments which led him to question whether Westphal's results were not slightly low owing to the use of lamp-black.

This and the preceding measurements of the coefficient of total radiation are excellent variations from the direct method. They are likely to give slightly low values, just as the thermometric methods, just described, give high values. Hence, these two methods serve the purpose of establishing upper and lower limits of the radiation constant.

(iv.) *Deductions of Lewis and Adams* (57).—In concluding the survey of what the writer considers the most reliable experimental determinations of the coefficient of total radiation, it is of interest to include in this paper a theoretical computation by Lewis and Adams (57) based upon data on the elementary electric charge e, the gas constant R, and the Faraday equivalent F. Their calculations lead to a value of $\sigma = 5{\cdot}7$.

The foregoing data are assembled in Table I., which gives also conservative corrections for atmospheric absorption, which is an important factor that in many cases was neglected by experimenters. By applying this obviously necessary correction it is interesting to find that these data, which were obtained by different methods, and which appear so discordant on first perusal of the original papers, can be brought into excellent agreement. They range about the number $\sigma = 5{\cdot}7$. In fact, two-thirds of the total number (12) of determinations recorded in Table I. are close to $\sigma = 5{\cdot}7$. The mean value of all these data which are free from question is $\sigma = 5{\cdot}72$ to $5{\cdot}73$.

If we neglected Todd's low value, determined from gas conduction experiments, and Puccianti's value, which is no doubt too high, the mean value remains unchanged. As we shall see presently, experimental evidence on ionisation potential, X-rays, and photoelectric work show that the value of the coefficient of total radiation is of the order $\sigma = 5{\cdot}7$.

TABLE I

OBSERVED VALUE AND THE MOST PROBABLE VALUE OF THE CONSTANT OF TOTAL RADIATION AFTER CORRECTING FOR REFLECTION, ATMOSPHERIC ABSORPTION, ETC.

Observer.	Date.	$\sigma \times 10^5$ erg.	Probable Value of $\sigma \times 10^5$.	Method.
Kurlbaum	1898	5·45	?	Bolometer
Féry	1909	6·3	?	Thermometer
Bauer and Moulin	1909	5·30	5·7	Thermopile
	1910	5·7	5·7	Pyrheliometer
Todd	1909	5·48	5·48	Gas conduction experiments
Valentiner	1910	5·58	5·69 to 5·75	Bolometer
Féry and Drecq	1911	6·51	?	Thermometer
	1912	6·2 5·57	5·68	Calibrated pyrometer
Shakespear	1912	5·67	5·67	Ratio of emissivities, metal : black body
Gerlach	1916	5·85	..	Modified Ångström pyrheliometer
	1920	5·80	5·80	
Puccianti	1912	5·96	5·96	Bolometer
	..	6·15	?	Thermometer
Westphal	1916	5·67	5·67	Ratio of emissivities, metal ; black body
Keene	1913	5·89	5·89	Thermometer
Coblentz	1915	5·72	5·72	Modified Ångström pyrheliometer
Kahanowicz	1919	5·61	5·69-5·73	Modified Ångström pyrheliometer

Mean value, $\sigma = 5{\cdot}72\text{-}5{\cdot}73 \times 10^{-5}$ erg cm.$^{-2}$ sec.$^{-1}$ deg.$^{-4}$.

III. THE CONSTANT OF SPECTRAL RADIATION

In order to appreciate fully the significance of the constant of spectral radiation, it is of interest to consider briefly the instruments and methods of observation.

§ (6) THE SPECTRORADIOMETER.—The determination of the constant, C_2, of spectral radiation requires very sensitive radiometric apparatus for measuring the intensities of the emission in different parts of the spectrum. In order to eliminate absorption, and especially the lack of achromatism which exists in lenses, the spectrometer is provided with mirrors as illustrated in *Fig.* 10, which shows the general arrangement of the apparatus required for such investigations. The radiations emanating from the black body A are focussed upon the spectrometer slit b by means of silvered mirrors placed in an air-tight box B. The spectrum is produced by means of a fluorite or quartz prism p (*Fig.* 10).

The radiometer for measuring the spectral intensities is shown at D, which in this case is a vacuum bolometer. The thermocouple for measuring the temperature of the furnace A is shown at Th. A sensitive ironclad Thomson galvanometer is used in connection with the spectrobolometer or thermopile.

The prism is calibrated by computing the minimum deviation settings, using for this purpose the refractive indices of fluorite (67) or quartz and the angle of the prism. As a fiducial mark or "zero" of the spectrometer circle the yellow helium line (15) is a convenient standard reference line.

By means of apparatus of this type the distribution of energy is measured in different parts of the spectrum. Two types of measurements may be made, viz. "Isochromatics" and "Isothermals."

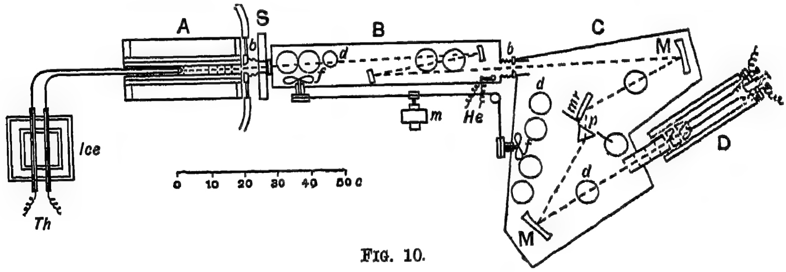


FIG. 10.

In making *isochromatic measurements* the spectrometer is set on a given wave-length, and the intensities are measured as the temperature of the radiator is varied.

In making *isothermal measurements* the temperature of the furnace is kept constant and the thermal radiation intensities are measured in different parts of the spectrum. This gives the prismatic spectral energy distribution[1] illustrated by the dotted curve (*Fig.* 11). On reducing these prismatic measurements to the standard normal spectrum (15) one obtains an energy distribution as illustrated in the continuous curve (*Fig.* 11). With rise in temperature the maximum of the emission curve shifts towards the short wave-lengths. This fact enters into spectral radiation formulas to be discussed presently.

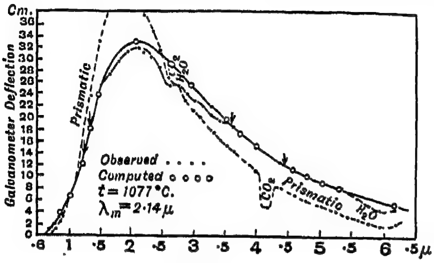


FIG. 11.

One of the difficulties encountered is the absorption by atmospheric water vapour and carbon dioxide, which produce depressions in the spectral energy curves as indicated in *Fig.* 11. However, by enclosing the spectrometer, by absorbing the moisture with phosphorus pentoxide, and by making the observations when the humidity is low, one can obtain trustworthy measurements over a wide region of the spectrum.

§ (7) FORMULAS FOR REDUCING SPECTRAL RADIATION DATA.—Planck's equation (2) is no doubt the closest representation yet formulated of the observed energy distribution in the spectrum of a black body. In *Fig.* 11 the circles represent the spectral energy distribution computed (66) by means of the Planck formula. In the spectral region, free from atmospheric absorption, to 6 μ, the observed and the computed curves agree to within the errors of observation, viz. 0·5 to 1 per cent.

Rubens and Kurlbaum (68) have made isochromatic observations, using the residual rays of fluorite in the spectral region of 24 μ and 52 μ. Their data also are in exact agreement with those computed by the Planck equation. From all of these data it appears that the Planck equation may be considered as representing, within the errors of observation, the energy distribution of a black body in the spectrum extending from 0·5 μ to 50 μ.

[1] See also "Radiation Theory," *Fig.* 1.

The Planck formula is based on theoretical considerations, and after almost two decades of discussion it remains unchanged. Recently, it is true, Nernst and Wulf (69) have arbitrarily modified the coefficient in the equation, changing C_1 to $C_1(1+\alpha)$, where α is a variable. Their whole procedure is based upon the assumption that the numerical value of the constant is $C_2=14300$ micron degrees, as observed from isochromatic observations, and that since the older values of C_2, obtained from isothermals, are somewhat higher, they must be reduced by a factor $(1-\alpha)$ depending upon the temperament of the radiator. They point out explicitly that the whole procedure depends upon the truth or falsity of the value of $C_2=14300$; or rather, as will be seen presently (see Table II.), upon the truth or falsity of the older values of $C_2=14400$ (about) obtained from isothermal measurements. Their deductions lead to a value of $\sigma=6{\cdot}04$, which, from a consideration of the whole subject on a subsequent page, appears to be much higher than the true value. Until we have more reliable experimental data from isothermal measurements, it does not appear necessary to consider their modification of Planck's equation.

(i.) *The Spectral Radiation Formulas—Isochromatic.*—After having made the spectral energy measurements, as indicated in the preceding pages, the next step is to compute the constant C_2 from the Planck formula (equation (2)). In view of the fact that some of the outstanding differences arise from methods of computing the constants from the experimental data it is important to consider the various reduction formulas which may be used in the computation.

For computing the constant C_2 from an *isochromatic* energy curve at any wave-length, λ, Planck's equation is used in the following form (15):

$$C_2=\frac{(\log E_2-\log E_1)\lambda T_1T_2}{\log e(T_2-T_1)}-\frac{(e^{-C_2/\lambda T_2}-e^{-C_2/\lambda T_1})\lambda T_1T_2}{T_2-T_1}, \quad (9)$$

where E_1 and E_2 refer to the emissivities corresponding to the absolute temperatures T_1 and T_2 respectively. In this equation the terms $\log(1-e^{-C_2/\lambda T_1})$, etc., which are similar to the terms in equation (13), are expanded in series, and only the first term of the expansion $(e^{-C_2/\lambda T_1}\log e)$ is used. An approximate value of $C_2=14300$ is used in applying the second term correction. For wave-lengths up to 1 μ this correction term

(to the Wien equation) is small, being only 2·1 and 4·3 for temperature intervals of $(T_2 - T_1)$ 363° and 623° respectively, when using $T_1 = 1450$. However, these corrections increase very rapidly with wave-length beyond 1 μ, so that at 2 μ with the same temperature intervals ($T_1 = 1450$; $T_2 = 1813$ and 2073°) the corrections to the values of C_2 amount to 168 and 227 respectively. This explains the rise in the value of C_2 with wave-length as found by Lummer and Pringsheim (61), who did not use the second term correction. Energy curves which coincide closely with the Wien equation give uniform values of C_2 with increase in wave-length when computed by using only the first term in equation as was found by Paschen. In their preliminary results Warburg (70) and his assistants, using a quartz prism, found that the value of C_2, obtained from isochromatics, is independent of the wave-length. On the other hand, they found an increase in C_2 with wave-length when using fluorite prisms.

(ii.) *Isothermal.*—Using an *isothermal* spectral energy curve, the constant of spectral radiation, C_2 (equation (2)), may be derived from the value of the wave-length of maximum emission, λ_m, by means of the equation

$$C_2 = a\lambda_m T, \quad . \quad . \qquad (10)$$

in which the value of a is 4·9651. The equation makes use of the Wien displacement law, which is the mathematical expression for the shifting of the wave-length of the maximum emission toward the shorter wave-lengths, with rise in temperature, viz.

$$\lambda_m T = A \text{ (a constant).} \qquad (11)$$

The value of the constant is close to $A = 2885$ micron degrees.

Experimenters are therefore concerned with the proper method of computing the position of the wave-length of maximum emission λ_m from the isothermal spectral energy curve, such as illustrated in *Fig.* 11.

In case one cannot observe the complete energy curve, as, for example, when using a quartz prism which absorbs strongly beyond 2·2 μ, it is possible to compute the value of λ_m from Planck's equation by reducing the observed energy curve to the normal spectrum, and taking from the energy curve the values of E_1 at λ_1 and E_m. The proper equation is

$$\frac{E}{E_m} = \left(\frac{\lambda_m}{\lambda}\right)^5 \cdot \frac{(e^{4 \cdot 9651} - 1)}{(e^{4 \cdot 9651\lambda_m/\lambda} - 1)}. \quad . \qquad (12)$$

This equation, which has been used extensively by Warburg (80) and his collaborators, appears to be well adapted for reducing the observations obtained with a quartz prism; but the temperatures must be sufficiently high so that the E_m does not fall in the region beyond 2·2 μ, where, as already mentioned, quartz begins to absorb strongly. This equation is not well adapted for reducing measurements made with a fluorite prism, owing to the small dispersion of the prism and hence the steepness of the energy curve on the short wave-length side, where one would want to use it, and also owing to the fact that, at the temperatures usually employed, the maximum emission E_m falls in the region of atmospheric absorption bands or in the region of 1·6 μ, where the dispersion curve of fluorite has a point of inflection. The latter point makes it quite difficult to determine accurately the factors for reducing the observations from the prismatic to the normal spectrum.

Another method of determining the values of λ_m from an isothermal spectral energy curve is by the "method of equal ordinates." For this purpose the normal spectral energy curve is drawn on accurately ruled co-ordinate paper, as illustrated in *Fig.* 11. Then, as indicated by the arrows, at two points where the intensities (the galvanometer deflections) are equal, $E_1 = E_2$, the corresponding wave-lengths λ_1 and λ_2 are read from the curve. The value of λ_m is easily derived (15) from equation (2) by equating $E_{\lambda_1} = E_{\lambda_2}$, and for the complete solution it is

$$\lambda_m = \frac{5(\log\lambda_2 - \log\lambda_1)\lambda_1\lambda_2}{a'(\lambda_2 - \lambda_1)\log e}$$

$$- \frac{\lambda_1\lambda_2[\log(1 - e^{-C_2/\lambda_1 T}) - \log(1 - e^{-C_2/\lambda_2 T})]}{a'(\lambda_2 - \lambda_1)\log e}. \qquad (13)$$

The second term in this equation can usually be abbreviated, since terms involving λ_1 are usually negligible. For values of λ_2 which are less than 4 μ the term $\log(1 - e^{-C_2/\lambda_2 T})$ may be expanded into a series and (by dropping all terms but the first) may be used in the form $-e^{-C_2/\lambda_2 T} \log e$. In this equation $a' = 4 \cdot 9651$. It may be noted that an approximate value of C_2 is necessary in order to obtain a solution for the second term factors in equation (13). This may be obtained by solving for λ_m by using only the first term in equation (13), which is the solution of the Wien equation as used by Paschen (59). It is sufficiently accurate (15) to use $C_2 = 14300$ in making this computation. A decrease of 100 units in C_2 (say $C_2 = 14200$) decreases the mean value of λ_m by 0·0012 μ.

The method of equal ordinates has been extensively used by Paschen (59) and by Coblentz (15, 66). It has several commendable features, because it is possible to utilise values of λ_1 and λ_2 which correspond with values of E_{λ_1} and E_{λ_2}, which are closely the same in magnitude as originally observed,

and hence contain the same errors of observation; also the slit widths are closely the same. This method does not require the most accurate wave-lengths and refractive indices (at least not for values of λ_2 greater than 4·5 μ), and it permits the selection of parts of the spectral energy curves which are free from atmospheric absorption bands. At a temperature of 1000° C. and at $\lambda_2 = 4·5\ \mu$ to 6 μ an error of $\lambda_2 = 0·05\ \mu$ changes λ_m by 0·007 μ to 0·008 μ. This amounts to 0·7 per cent in the value of C_2, and is much greater than would occur in practice.

The method of equal ordinates necessitates reducing the prismatic spectral energy measurements to a normal spectrum, plotting the data upon co-ordinate paper, and drawing in a smooth curve from which the values of λ_1 and λ_2 are read, corresponding with the equal ordinates $E_{\lambda_1} = E_{\lambda_2}$. Although it requires but little additional time to plot the data, after having made the observations, if one is certain that the observations lie close to the curve, the obviously logical procedure is to compute the spectral radiation constants from any two observed points E_{λ_1} and E_{λ_2}. This, however, does not shorten the observational work, for it will be necessary to make observations at three adjacent spectrometer settings (separated by the width of the radiometer receiver) in order to determine the spectral purity factor which is required in order to reduce the data to a normal spectrum. After spending months in obtaining the data, the question of saving time in their reduction is of minor importance. Furthermore, it is desirable to draw the complete curve and preserve it for future reference. If the observations do not lie close to the smooth curve, then this method is slightly arbitrary as to the manner of combining the observations so as to obtain an integration of the whole curve without making more computations than would be used in the "method of equal ordinates."

The appropriate formula for computing the radiation constants from any two observed points, E_{λ_1} and E_{λ_2}, on an isothermal spectral energy curve is easily deduced from Planck's equation, and in its complete form, as shown by Dellinger (71), is

$$C_2 = \frac{\lambda_1\lambda_2 T}{(\lambda_2 - \lambda_1)\log e}\left[\log\frac{E_{\lambda_2}}{E_{\lambda_1}} + 5\log\frac{\lambda_2}{\lambda_1} - \log\left(\frac{1 - e^{-C_2/\lambda_1 T}}{1 - e^{-C_2/\lambda_2 T}}\right)\right]. \quad (14)$$

The term $\log(1 - e^{-C_2/\lambda T})$ may be expanded into a series, and usually all the factors can be neglected, except one which is $\log e \,.\, e^{-C_2/\lambda_2 T}$. It is therefore necessary to know the approximate value of C_2, as is required in the method of computation used by Coblentz (15, 66).

In concluding this discussion we may notice a calculation of the constant C_2 by an extension of the theory of least squares, by Roeser (72). Using the data illustrated in *Fig.* 11, his calculation gave a value of $C_2 = 14342$, as compared with $C_2 = 14339$ computed by Coblentz (66) (*loc. cit.* p. 462), by the method of equal ordinates.

(iii.) *Spectrophotometric.*—A spectrophotometric method may also be used in determining the constant C_2. It consists in determining the ratio of brightness of a black body at, say, the melting-point of gold (1063° C.) and at some higher temperature, say the melting-point of palladium (about 1556° C.). This requires a knowledge of the wave-length ($\lambda = 0·65\ \mu$ is usually taken) of the light photometered and the absolute temperatures T_1 and T_2 of the radiator. The appropriate formula is

$$\log\frac{E_2}{E_1} = C_2 \log e\left(\frac{1}{T_1} - \frac{1}{T_2}\right). \quad (15)$$

The intensities E_1 and E_2 are usually measured with an optical pyrometer. If the constant C_2 is known the temperature T_2 (say 2000° C.) may be computed. Conversely, using known temperatures T_1 and T_2 the constant may be determined. As a matter of fact the method has never been used very successfully in a quantitative manner because, as will be noticed presently, the higher temperature T_2, at some fixed temperature point (say the melting-point of palladium), is not accurately known. Recent experimenters are therefore proceeding in the reverse order, and assume that if the constant $C_2 = 14300$, then the melting-point of palladium is 1557° C., or if $C_2 = 14350$, then the melting-point of palladium is 1550° C. The task before us is therefore to obtain an estimate of the probable value of the constant C_2.

§ (8) Determinations of Constant of Spectral Radiation.—In spite of all that has been published on the partition of energy in the spectrum of a black body there are but few experimental data at hand which are more than qualitative in value.

To the writer it seems that all these early data should be considered from the standpoint of historical interest, since it seems impossible to give them much weight in connection with the results obtainable at the present day.

(i.) *Observations of Paschen* (59, 60).—The pioneering work in spectral radiation measurements and constants was inaugurated by Paschen (59), who observed an extensive series of spectral energy curves at temperatures ranging from 100° to 450° C., using six different kinds of bolometers covered with different kinds of absorbing material, *e.g.* lamp black and platinum black or copper oxide; also having

the bolometers in the focus of a hemispherical mirror. The radiators were heated by means of boiling water, aniline, sulphur, etc. The mean value of all his observations, which are in close agreement, was $A=2891$ and $C_2=14352$. He continued the work at temperatures ranging from 400° to 1050° C., using a large porcelain radiator. He used also metal cups of copper or platinum, blackened with oxides of copper or iron, which were heated within this porcelain radiator; making in all about a dozen different arrangements of the radiators. The bolometer was covered with platinum black and was situated in the focus of a hemispherical mirror to "blacken" it. The mean of the new series was $A=2907$, with a probable error of ± 16.

In a further investigation Paschen (60) undertook the work anew, after redetermining the reflecting power of silver and the refractive indices of fluorite. He used a porcelain tube radiator, also three other radiators, which he blackened as in the preceding research. He took the wise precaution to project an image of the radiating wall of the black body upon the spectrometer slit (see *Fig.* 10), in order to avoid possible radiations from the side walls and diaphragms falling upon the prism and bolometer. He made complete corrections for the selective reflection from the prism. He then found that his observations fitted neither the Wien nor the Planck equation, the values on the long wave-length side of the energy curves falling between the two theoretical curves. He found that if (for reasons he himself could not explain, *loc. cit.* p. 295) he multiplied his observations by factors varying from 1·02 at 3·91 μ to 1·195 at 8·25 μ, etc., the observed energy curve would fit the Planck equation and fulfil better the condition of congruence. Upon this basis, and presumably by computing the λ_m by using the first part of equation (13), he obtained a value of $A=2921$ for his newest data. He then applied similar factors to some of his previous data, thus making them agree better with the Planck equation, and the value of A was increased from 2891 to 2915, or about 0·87 per cent. From this it appears that, if he had not multiplied his observations by these arbitrary factors, his latest results would have been about 0·87 per cent lower or $A=2894$, which is practically the same as the value previously obtained.

Coblentz (15, 66) obtained a plausible estimate of the correction to Paschen's values by applying the values of the second term which result from computing λ_m by equation (13). This is admissible, because Paschen's observed curves are said to fit the Planck equation to about 4 μ. The conclusion arrived at (*loc. cit.* (66), p. 468) is that Paschen's recent determination is of the order $A=2894$. This is close to the earlier determination $A=2891$, which is probably more reliable as regards the temperature scale. In other words, Paschen's original data fall within the range of the recent and more accurately determined values, being of the order $C_2=14360$ micron degrees.

(ii.) *Observations of Lummer and Pringsheim* (61, 63, 74).—Their spectral radiation measurements were made on porcelain tube radiators, illustrated in *Fig.* 1. The radiator was placed directly in front of the spectrometer slit. This reduces the length of air-path, and hence the absorption; but there is uncertainty in keeping the alignment, especially since in their designs no attempt was made to prevent the tube from sagging.

As is true of all the older measurements on radiation, operated above 1200° C., the temperature scale is defective (too high), giving values of C_2 which are too high.

The values of the constants published by Lummer and Pringsheim (61) require a treatment similar to that just given to Paschen's data. They used the earlier indices of refraction of fluorite, published by Paschen (73) in 1894, which are in error in wave-length by 0·02 μ for the region of 1 μ to 2·5 μ. Fortunately most of their values of λ_m are greater than 2·5 μ and no correction is required. But little information is at hand as to how they calculated their values of λ_m. In their earlier work (61) they say that, after bridging over the absorption bands by means of a smooth line, the values of λ_m and E_m may be read directly from the spectral energy curve. This would permit the determination of only one reading of λ_m, whereas the method of equal ordinates permits the calculation of a number of values of λ_m.

In their first investigation a series of four spectral energy curves, observed at temperatures of 837° to 1416° abs., gave a value of $A=2879$.

For a second series of measurements they enclosed the spectrometer in order to dry the air. A series of five spectral energy curves, observed between the temperatures of 814° and 1426° abs., gave a value of $A=2876$. These two values are in close agreement with Paschen's determinations made at that time. Since they observed the value of λ_m at E_m, instead of by the method of equal ordinates, the second term correction of equation (13) does not enter into the calculation. They found that their data did not fit the Wien equation, and this no doubt gave impetus to the formulation of the Planck equation. Suffice it to say that their data, obtained up to this time, if reduced on the basis of the Planck factor $\alpha=4·9651$, lead to a value of $C_2=14290$. In a subsequent investigation (63) the radiator was heated to a much higher

temperature — to 1646° abs. A series of 8 isothermal spectral energy curves gave a value of $A=2940$ and $C_2=14590$. In view of the fact that a correction of about 0·02 μ must be made for the calibration at 1 μ to 2·5 μ, this would reduce the value of λ_m and of C_2 by 2 per cent, or $C_2=14300$. Although the apparatus was enclosed, in order to remove the CO_2 and water vapour, the energy curves are appreciably affected by atmospheric absorption.

As in previous work, their temperature scale above 1000° C. was obtained by extrapolation. Just how much the thermocouple calibration may be in error is not recorded. In a subsequent paper (75) they revised the temperature scale used in the previous test (9) of the Stefan-Boltzmann 4th-power law. The revised temperatures are 10 to 12° lower, and, at the highest temperatures, they are 20 to 25° lower than previously used. Hence the value of A (see equation (11)) would be lowered. Whether any temperature correction of this magnitude must be made to the spectral radiation data is not stated; though the researches on the gas thermometer by Holborn and Day (76) were then in progress. Suffice it to say that although their later value of $A=2940$ is the one frequently quoted in books, none of the subsequent investigators have found a value of A which is within 15 per cent as high as is this one.

From the foregoing consideration of the data obtained by Lummer and Pringsheim it appears conservative to place their value at $C_2=14300$ micron degrees.

(iii.) *Observations of Warburg and Collaborators* (80 to 90).—During the past twenty years the determination of the constants of radiation and their application to the temperature scale has been relentlessly pursued by the Phys. Tech. Reichsanstalt (75) at Berlin.

During the past ten years these investigations have been carried out by Warburg and his collaborators. They used fluorite and quartz prisms and vacuum radiators. The sodium line was used as a zero setting of the spectrometer.

The radiometer was a vacuum bolometer, operated by a null method. In this manner the galvanometer acted merely as an indicator. The temperature fixed points were the melting-point of gold and some higher temperature, *e.g.* melting-point of palladium.

In their first communication (80, 81) they reported a value of $C_2=14570$ on the basis that the melting-point of palladium is 1549° C. On the other hand, if the higher temperature point was determined radiometrically, by extrapolation from the gold point, using the Stefan-Boltzmann law, then the value of C_2 was found to be the smaller. Hence they questioned the gas thermometer temperature scale and proceeded to make their investigations at high temperatures by using the radiation laws to establish their scale of temperatures. In this manner they avoided the temperature scale as transferred from the gas thermometer by means of thermocouples. They retained only one temperature fixed point, viz. gold at 1064° C. (later 1063°). The higher temperatures (1400° C.) were determined radiometrically. For this purpose they observed the position, E_m, of the isothermal spectral energy curve reducing the data by means of formulas (12) and (10).

Using Paschen's refractive indices of refraction of fluorite and Carvallo's (93) indices of quartz, in their next communication (82) they report values which varied from $C_2=14200$ to 14600.

Using improved methods for adjusting the sodium lines on the bolometer, and making further provision so that only radiation emanating from the central diaphragm of the radiator was incident upon the bolometer, in their next report (83) the fluorite prism gave values ranging from $C_2=14300$ to 14600, and it was discarded. A quartz prism gave a value of $C_2=14360$ micron degrees.

In a very complete investigation (84, 85) they repeated the previous work. Using a quartz prism the values obtained are $C_2=14370\pm40$ and $A=2894\pm8$ for the temperature interval of 1337° and 2238° abs.

Their next step was to cut a prism out of a block of quartz of which the absorptivity had been determined previously. With this and other improvements, including different radiators, a value of $C_2=14250$ was obtained (87, 88), from the temperature scale based on the Stefan-Boltzmann law of total radiation; and a value of $C_2=14300$ or 14400 was obtained by using the Wien displacement law (*e.g.* 11) to establish the temperature. The uncertainty in the value of $C_2=14300$ or 14400 is owing to the uncertainty in their calibration curve (refractive indices) of the quartz prism.

The present position of their work has led them to the adoption (89) of the value of $C_2=14300$ and the melting-point of palladium $=1557°$ C. Subsequent investigations (90) appear to be made on this basis.

There is an uncertainty of perhaps 1° in the temperature scale at the melting-point of palladium (92). In view of the great variations in the various determinations of C_2, it was perhaps a wise decision for Warburg to adopt the round number $C_2=14300$ micron degrees for the spectral radiation constant; though, as we shall see presently, theory and other experimental data would place the value somewhat higher.

(iv.) *Observations of Coblentz* (15, 39, 66).—In this investigation the spectrum was produced by means of a mirror spectrometer and a fluorite prism, as illustrated in *Fig.* 10. In the first work an air bolometer, and later a vacuum

bolometer, was used for measuring the partition of energy in the spectrum. The radiator (*Fig.* 1) was a porcelain tube, wound with platinum ribbon, through which electric current was passed.

It was operated at temperatures ranging from 450° C. to 1500° C. Various incidental questions, such as the adjustments of the optical parts of the apparatus, the temperature uniformity and temperature control of radiator, the water-cooled shutters, the temperature scale, the method of reduction of the observations to the normal spectrum, the proper formulæ for computing the numerical constants, etc., were investigated.

The first paper (15) contains also data on various subsidiary problems such as (1) the variation of the reflecting power of silver with angle of incidence and with wave-length, (2) the variation of reflecting power of fluorite with angle of incidence and with wave-length (refractive index), and (3) data for reducing the observations from prismatic to normal spectrum.

As already explained in the discussion of the methods of reducing the observations, it was decided to observe isothermal spectral energy curves as illustrated in *Fig.* 11, and compute the position of λ_m by the method of equal ordinates, using equation (13).

From the very beginning of this investigation on black body radiation it was found that the Wien equation did not fit the spectral energy curves. The assumption was therefore tentatively made that the observed curves fit the Planck equation; and at the completion of the investigation this was found to be true for about 75 per cent of all the observed curves. This conclusion was based upon the uniformity of the values of λ_m which resulted from computation (by the method of equal ordinates $E_{\lambda_1}=E_{\lambda_2}$) of values of λ_1 and λ_2 which were taken far apart, and also close together, on the observed isothermal spectral energy curve.

The observations were made in the winter when the humidity was low, and the investigation extended over four winter seasons. An attempt was made to obtain a great many isothermal spectral energy curves, so as to avoid the personal bias which can enter the working over of a few curves. This was probably a mistake; for no attempt could be made to correct the observations for small changes in temperature and bolometer sensitivity.

In the meantime, owing to impairment of eyesight, the reduction of data had to be entrusted to others who were not familiar with the work. The first calculation of the data gave a value of about $C_2=14350$. But doubts arose concerning the calibration curve of the prism. A new calibration curve was worked out and the data recalculated and published as being $C_2=14456$.

Data were obtained also with a fluorite prism which was full of cleavage planes. This produced much scattered radiation which distorted the energy curves at 4 μ to 5 μ. These data were therefore not used in the calculations.

In the meantime Paschen (67) published further data on the dispersion of fluorite which indicated that the calibration curve, and hence the value of $C_2=14456$, was wrong.

In the earlier work the temperature scale was also in doubt. The last series (1912) was not observed at temperatures much above 1200° C., and hence is quite free from doubts about the temperature scale, which was 1° lower than previously used.

The second paper (66) on this subject dealt with a recalculation of these data, using a revised calibration curve. The mean value of the spectral radiation constant based on 93 spectral energy curves (series of 1912) is $C_2=14353$.

If the corrections to the temperature scale (mentioned in the previous paper (15)) are applied, the value is $C_2=14362$. A further correction $(=+7)$ is necessary because the second term in equation (13) was computed, using $C_2=14300$ instead of $C_2=14350$. Hence the final corrected value, as published in the second paper, is $C_2=14369$ micron degrees and $A=2894$ micron degrees.

In order to obtain a check on the method of calculation of the constant a least square reduction of the first isothermal curve of the series of 1912 was undertaken by Roeser (72). He obtained a value of $C_2=14342$, which is in agreement with the value computed by the method of equal ordinates, viz. $C_2=14339$.

Recently a further examination (39) was made of the accuracy of the factors used in converting the previously observed (15) prismatic spectral energy data into the normal energy distribution. The graphical methods previously employed were checked and similar factors were obtained by computation, using the first differential of the dispersion formula, which best represents the observed refractive indices of fluorite. These refractive indices were obtained from consideration of all the available data, which, in the region of 1 μ to 2 μ, are represented by the curve published by Langley and Abbot (94). The best dispersion formula is that of Paschen (97). However, owing to incompleteness of the formula, the graphical method was found to be just as accurate as was the method of computation.

The conclusion (39) arrived at was that the spectral radiation constant $C_2=14353$ micron degrees, determined some years ago (66), remains unchanged. However, at that date there was some doubt as to whether some of the corrections then applied should have been made, giving a value of $C_2=14369$—*i.e.* the value might be $C_2=14366$.

Rather curiously and unfortunately in all

these inquiries into the small errors that eight years ago were considered negligible the one concerning the zero setting has remained unconsidered until now. The calculation of the calibration curve was based on the sodium lines, $\lambda = 0{\cdot}5893\ \mu$, for a reference point. Subsequently Coblentz adopted the then novel procedure of using the yellow helium line, $\lambda = 0{\cdot}5876\ \mu$, instead of the sodium lines for adjusting the zero setting of the bolometer. For this purpose the refractive index $n = 1{\cdot}43390$ was adopted. Subsequently Paschen (67) published the value of $n = 1{\cdot}433907$ ($17{\cdot}5°$ C.) for the refractive index of the helium line, and $n = 1{\cdot}433866$ for the sodium lines. He observed a difference of 12″ in the minimum deviation settings of the sodium and the helium lines.

On the basis of Paschen's value of the refractive index of fluorite for the yellow helium line, $n = 1{\cdot}433877$ (at 20° C.), there is a difference of 6″ between the computed and the observed zero setting of the spectrometer circle. As a result, the average values of λ_m, in terms of the spectrometer circle, must be reduced by 6″. Since the wave-lengths λ_m occur between 2 μ and 3 μ (and the majority at about $2{\cdot}2\ \mu$), the 6″ are equivalent to $0{\cdot}003\ \mu$ to $0{\cdot}004\ \mu$ in this part of the spectrum. From equation (11) it may be noticed that this amounts to a reduction of the previously published value of the constant $C_2 = 14369$ by 0·3 to 0·4 per cent. This gives a value of $C_2 = 14311$ to 14326 micron degrees.

As already stated, the second paper (66) was the result primarily of a revision of the calibration curve of the fluorite prism used; and it is unfortunate that at that time the above application of the temperature coefficient of the refractive index of the helium line was overlooked.

The citations of the foregoing researches suffice to show some of the difficulties under which experimenters are labouring. One set is concerned with the temperature scale; another is determining the optical constants of the prism material; and a third group, using both the temperature scale and the optical constants, is engaged on the radiation constants. In turn, the first set of experimenters must apply the results of those working on the radiation constant in order to verify and extrapolate the temperature scale. In the meantime the second experimenter improves his measurements of the optical constants (refractive indices) of the prism material, while a fourth enters the field and adds refinements by determining the temperature coefficient of the refractive indices. Then the unpleasant task arises to recalculate the prism calibration and the numerical value of the spectral radiation constants. Added to these difficulties is the constant change in personnel, as is apparent from perusal of the title-pages of the published papers.

Fortunately the observations are in terms of the spectrometer circle scale, and, if the necessity arises, as it did in the case just discussed, one can revise the corresponding wave-lengths and recalculate the constants.

The various determinations of the constant of spectral radiation C_2 are assembled in Table II. The sixth column gives the probable value as determined from consideration of the data in the text. The mean value is $C_2 = 14320$ micron degrees. The latest and most reliable determinations of the national laboratories are close to this value. Unfortunately, perhaps, the average value is so close to the theoretical value, arrived at from consideration of photo-electrical and similarly related phenomena, that experimenters may be led to consider their task finished instead of just begun.

§ (9) Optical Pyrometer Measurements.—Various attempts have been made to determine the constant C_2 by means of an optical pyrometer, using equation (15). There are many difficulties to be overcome before one can conclude that the data so obtained are trustworthy. One of the uncertainties is the effective wave-length of the red glass used in the eyepiece of the pyrometer. Other difficulties are encountered when using a spectral pyrometer. Also the temperature scale is in doubt. Hence recent experimenters have not attempted to determine C_2, but reverse the process and, working on the assumption that C_2 is a certain value, say $C_2 = 14350$, determine the melting-point of palladium. This seems to be the preferable procedure.

(i.) *Observations of Holborn and Valentiner* (64, 65).—In 1906 Holborn and Valentiner (64) obtained a value of $C_2 = 14200$ as a result of a series of spectrophotometric measurements using formula (16). This is now admittedly in error, owing to an erroneous temperature scale in which the melting-point of palladium was taken to be 1575° instead of about 1556°, which would give a higher value of C_2. Recently Valentiner (65) corrected this value for lack of blackness of the radiator, etc., raising it to $C_2 = 14350$. This is one of the difficulties and uncertainties experienced by experimenters who have attempted to determine the constant C_2 by methods requiring a temperature scale which is higher than about 1200° C.

(ii.) *Observations of Mendenhall* (78).—One of the few recent direct determinations of C_2 by optical pyrometer methods was made by Mendenhall (78). In establishing a temperature scale he took the Day and Sosman value

of the melting-point of palladium, viz. 1549° C. From the ratio of brightness of the black body at the melting-point of gold and palladium he obtained the value of C_2 = 14413, as a check on his measurements. Correcting for the low temperature scale, which amounts to 3·6 to 5° C. on the basis that the melting-point of Pd is 1555° to 1557°, the value of C_2 is decreased by 0·23 per cent to 0·31 per cent or C_2 = 14368 to 14380.

It is difficult to obtain high accuracy in optical pyrometer measurements on a radiator to be open to question. Just why the temperature T_1 was not similarly determined radiometrically is not clear. The temperature T_1 = 1604° abs. (1331° C.) seems to be in doubt in view of the present-day belief that the melting-point of palladium is 1555° C. to 1557° C.

If the melting-point of Pd is 6° to 8° higher than the value used in this determination, then the value of C_2 is 0·22 to 0·3 per cent higher than his published value (C_2 = 14392) and is of the order of C_2 = 14425 to C_2 = 14435.

Table II

Observed Value and the Probable Value of the Constant C_2 of Spectral Radiation

Observer.	Date.	$\lambda_m T = A$ Observed.	$\lambda_m T = A$ Corrected.	$C_2 = a\lambda_m T$ and from Isochromatics.	C_2 Probable Value.	Remarks.
Paschen . .	1899	2891	2891	..	..	Fluorite prism
	..	2970 ?	2907 ?	..	14360 }	Temperature scale is questioned
	1900	2921 ?	2894	..	.. }	
Lummer and Pringsheim	1900	2879	2879	14290	..	Fluorite prism. Wave-length calibration of prism is questioned; also temperature scale.
	..	2876	2876	..	..	
	..	2940	2882	14310	14300	
Warburg and Collaborators	1911	..	..	14200 to 14600	.. }	Fluorite prism
	1912	..	..	14300 to 14400	.. }	
	1912	..	..	14360	.. }	Quartz prism
	1913	2894	..	14370	.. }	
	1915	..	..	14250	..	Temperature from Stephan Boltzmann law
	1915	..	..	14300 to 14400	14300	Refractive indices (calibration of prism) questioned
Coblentz. .	1913	2911	..	14456	..	Fluorite prism
	1916	2894	..	14369	..	Revised calibration (refractive indices) of prism; 1912 data are recalculated
	1920	..	..	14311 to 14326	14318	Correction for zero setting of bolometer

Average value, C_2 = 14320 micron degrees.

at the melting-point of gold. In the main part of his research Mendenhall therefore established his lower fixed point T_1, as that having 14·91 times the intensity of radiation of a black body at the melting-point of gold, for the complex wave-lengths transmitted by a certain standard pyrometer glass.

Subsequently, by direct comparison with Day and Sosman's standard thermocouple, which was calibrated presumably on the basis that the melting-point of palladium is 1549° C., this temperature T_1 was found to be 1331° C.

The higher fixed point T_2 = 2705° abs. was determined radiometrically on the basis of the Stefan-Boltzmann law, and does not appear

(iii.) *Observations of Hoffman and Meissner* (79, 90).—These experimenters used an especially constructed black body in a bath of the molten metal. The ratio of the brightness of the black body at the melting-point of palladium is determined relative to the brightness of the same at the melting-point of gold. They found that if the melting-point of palladium is taken 1549° C. (scale of Day and Sosman), then C_2 = 14440 (79). In a further investigation (89) they found that if C_2 is 14300, and if the melting-point of gold is 1063°, then the melting-point of palladium is 1557° C. This is in agreement with the observations of Holst and Oosterhuis (91), who found a value of C_2 = 14465 if the melting-point of Pd is

1549°, and $C_2 = 14300$ if the melting-point is 1557° C.

(iv.) *Observations of Hyde and Forsythe* (92).—In their earlier experiments they found a value of $C_2 = 14460$ on the basis that palladium melts at 1549° C. The great inconsistencies in these and other data led them to use the value $C_2 = 14350$, which is commonly used by American experimenters.

Using a carefully constructed spectral pyrometer, and on the basis that the value of $C_2 = 14350$, their most recent measurements place the melting-point of palladium at 1555·5° C.

The outstanding question then is: What is the value of C_2? Is it $C_2 = 14300$ or 14350? If we accept 1556° C. as the melting-point of palladium, then Hoffmann and Meissner's data indicate a value of $C_2 = 14315$.

Both sets of experimentors (*i.e.* those interested in the radiation constants and those interested in the temperature scale) seem to realise the difficulty in attempting to define the value of the constant C_2 or the melting-point of palladium in terms of this constant. Hence as a compromise basis for future work and for future adjustments it seems appropriate to adopt the value of the spectral radiation constant $C_2 = 14320$ and the melting-point of palladium at 1556° C.

IV. Verification of the Laws of Radiation

In the foregoing pages an inquiry is made into the instruments and methods used in, and the numerical values of various determinations of, the constants of radiation.

The various methods are classified, and a brief description is given of each research. An attempt is made to indicate the good and the defective features in each research. This represents not only the writer's opinions but also those of other experimenters.

It is shown that the major part of the variation in the various determinations of the numerical values of the constants, especially the constant of total radiation, is owing to the fact that, in the original papers, no corrections were made for atmospheric absorption of radiation in its passage from the radiator to the receiver. Conservative corrections for atmospheric absorption are made to the various determinations in which such corrections had not been made. As a result, instead of having wide variations, there is a remarkably close agreement in the numerical values of the various determinations which are free from other obvious defects. Although it cannot be said that the true numerical values are exactly as here recorded, it is evident that the time is past when the value of the constants of radiation are swayed by a single and perhaps novel method of research. The best that an experimenter should expect is that his own little contribution to the subject may have sufficient merit to go into the melting pot with the other determinations.

§ (10) The Formula and the Coefficient of Total Radiation.—In the foregoing review the data are assembled and the evidence weighed *pro* and *con*. It is shown from various experiments that, beyond all reasonable doubt, the total radiation emitted from a uniformly heated enclosure is proportional to the 4th power of the absolute temperature—the so-called Stefan-Boltzmann law. Furthermore, the tabulated data show that the numerical values of the majority of the various determinations of the coefficient, σ, which enters into the mathematical formula of total radiation, range about the value given by the expression $\sigma = 5{\cdot}7 \times 10^{-5}$ erg cm.$^{-2}$ sec.$^{-1}$ deg.$^{-4}$. The average of 12 of the most reliable determinations is $\sigma = 5{\cdot}72$ to $5{\cdot}73 \times 10^{-5}$ erg cm.$^{-2}$ sec.$^{-1}$ deg.$^{-4}$.

§ (11) The Formula and the Constant of Spectral Radiation.—Experimental evidence is cited showing that throughout the spectrum from 0·5 μ to 50 μ Planck's formula fits the observed spectral energy distribution more closely than any other equation yet proposed. This formula is based upon theoretical principles, and after two decades of discussion it remains unchanged.

The constant C_2 which determines the slope of the spectral energy curve has been the subject of numerous investigations. The numerical value of C_2 has fluctuated considerably in the various determinations. In the foregoing pages it is shown that this is owing to experimental difficulties, such as, for example, lack of precise knowledge of the temperature scale, and of the refractive indices of the prisms used. The tabulated data show that the various determinations of the constant of spectral radiation are of the order of $C_2 = 14300$ to 14350, with a mean value of $C_2 = 14320$ micron degrees.

§ (12) Confirmatory Evidence.—One of the most interesting phases of the inquiry into the laws and constants of radiation is the confirmatory data which one obtains from a consideration of the inter-related phenomena of atomic structure, of X-rays, of ionisation and resonance potential, and of photo-electrical action. From these data, as well as from the foregoing data on the two constants of radiation, C_2 and σ, one can compute the value of Planck's element of action, h. This gives seven independent methods of determining the universal constant h. Or from any one of four of these methods one can calculate (99, 100, 101) the radiation constants; and it seems truly remarkable how close the calculated values agree with the observed values of the radiation constants.

For making these calculations from Planck's

radiation[1] theory (2) we have the following relations:

$$C_2 = chk^{-1} = 4{\cdot}9651\ \lambda_m T, \qquad (16)$$

$$\sigma = \frac{ac}{4} = \frac{12\pi \times 1{\cdot}0823k^4}{c^2h^3}, \qquad (17)$$

$$\lambda_m T = \frac{ch}{4{\cdot}9651k}, \qquad (18)$$

$$k = \frac{eR}{cF}, \qquad (19)$$

while the value of c_1 is $8\pi hc$.

In these equations the constants have the following values:

h (Planck's constant)	$=6{\cdot}55 \times 10^{-27}$ erg sec.
k (Boltzmann gas cons.)	$=1{\cdot}372 \times 10^{-16}$ erg deg.$^{-1}$.
c (Velocity of light)	$=2{\cdot}9986 \times 10^{10}$ cm. sec.$^{-1}$.
F (Faraday's constant)	$=96500$ coulombs.
R (Absolute gas constant)	$=831{\cdot}5$ erg deg.$^{-1}$.
e (Unit electric charge)	$=4{\cdot}774 \times 10^{-10}$ e.s.u.

From equation (19) it may be noted that a change in the value of e affects the value of C_2 directly, while the value of σ is affected by e^4.

The data computed from the above-mentioned constants and formulae are illustrated in *Fig.* 12, from which it is an easy matter

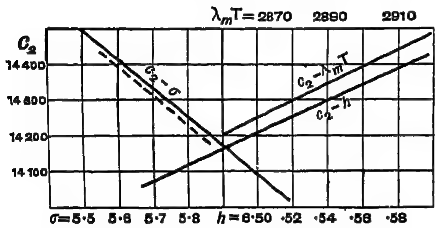


FIG. 12.

to compare experimental data (101). For example, Coblentz's (29) value of the coefficient of total radiation is $\sigma = (5{\cdot}72 \pm 0{\cdot}012) \times 10^{-5}$ erg cm.$^{-2}$ sec.$^{-1}$ deg.$^{-4}$. This indicates a value of $C_2 = 14321$ micron degrees (which is close to the average determination) and a value of $h = 6{\cdot}551 \times 10^{-27}$ erg sec. The value of h, determined by Blake and Duane (98), by X-rays, is $h = 6{\cdot}555 \times 10^{-27}$ erg sec.; or an indicated value of $C_2 = 14330$ micron degrees, which is close to the average of the direct experimental determinations of this constant.

Again, starting with Warburg's value of $C_2 = 14300$ the corresponding value of the coefficient of total radiation is $\sigma = 5{\cdot}74$, which is the higher estimated limit of the average of 12 different determinations of this constant. Hennig (99), on the basis of Sommerfeld's (102) theory, the measurement of spectral lines, and the value of the electron, obtains $C_2 = 14320$ and $\sigma = 5{\cdot}717$.

In summing up the evidence it is of interest to include Birge's (100) comprehensive and exact calculations of the constant h. In these calculations he, of course, assumes the truth of (1) Lewis and Adams' (57) theory of ultimate rational units; (2) of the relation between σ and h, as given by Planck's radiation formula; (3) of the quantum relation as applied to X-ray data (98); (4) of Einstein's photo-electric equation; (5) of Bohr's theory of atomic structure; and (6) of the quantum relation as applied to ionisation and resonance potentials.

In this manner he obtains seven separate calculations of Planck's constant of action h, the least square mean value of which, as shown in Table III., is $h = 6{\cdot}5543 \times 10^{-27}$ erg sec. This is close to the average of the value of h, which results from consideration of the two radiation constants.

TABLE III

BIRGE'S CALCULATION OF PLANCK'S UNIVERSAL CONSTANT h BY VARIOUS METHODS

Value of h.	Method.	Remarks.
$6{\cdot}551 + 0{\cdot}009$	$\sigma = 5{\cdot}72$	Total radiation
$6{\cdot}557 \pm 0{\cdot}013$	$C_2 = 14330$	Spectral radiation
$6{\cdot}542 \pm 0\ 011$	Rydberg constant	Bohr's theory of atomic structure
$6{\cdot}578 \pm 0{\cdot}026$	Photo-electric equations	Einstein's equation
$6{\cdot}555 \pm 0{\cdot}009$	X-rays	Quantum relation
$6{\cdot}560 \pm 0{\cdot}014$	Ultimate rational units	Theory of Lewis and Adams
$6{\cdot}579 \pm 0{\cdot}021$	Ionisation potential	..

Mean value, $h = (6{\cdot}5543 \pm 0{\cdot}0025)10^{-27}$ erg sec.

From this calculation and intercomparison by Birge (100) of the data on C_2, σ, and h, as determined by thermal radiometric, photo-electric, X-rays, and ionisation potential measurements, it appears that the value of h, computed from radiometric data, is in close agreement with that obtained by more direct measurement. In other words, it appears to prove the validity of laws of radiation and to establish the level of the numerical values of the constants entering therein.

The outstanding disagreement between all the observed and computed data appears to be of the order of 1 to 3 parts in 1000, whatever the method or experimentation. This is a very close agreement, considering the variety of the data and the difficulties involved in making the experiments. It seems to indicate

[1] See "Radiation Theory," § (6).

something more than a fortuitous relation between properties of matter.

In conclusion, it may be added that to a close degree of approximation we have the following constants:

Melting-point of palladium = 1556° C.

$C_2 = 14320$ micron degrees.

$\lambda_m T = 2885$ micron degrees.

$\sigma = 5{\cdot}72 \times 10^{-5}$ erg cm.$^{-2}$ sec.$^{-1}$ deg.$^{-4}$.

$h = 6{\cdot}55 \times 10^{-27}$ erg sec.

W. W. C.

V. Bibliography

1. Stefan, *Sitzungsber. d. K. Gesellsch. der Wissensch. zu Wien*, 1879, (2), lxxix. 391.
2. Boltzmann, *Wied. Ann. der Phys.*, 1884, xxii. 31 and 291.
3. Planck, *Vorlesungen über Wärmestrahlung*.
4. Draper, *Amer. Jour. Sci.*, 1847, (2), iv.; *Phil. Mag.*, 1847, (3), xxx.; *Scientific Memoirs*, London, 1878.
5. Kirchhoff, *Pogg. Ann. der Phys.*, 1860, cix. 275.
6. Christiansen, *Wied. Ann. der Phys.*, 1884, xxi. 364.
7. Saint-John, *Wied. Ann. der Phys.*, 1895, lvi. 433.
8. Wien und Lummer, *Wied. Ann. der Phys.*, 1895, lvi. 451.
9. Lummer und Pringsheim, *Wied. Ann. der Phys.*, 1897, lxiii. 395; *Ann. der Phys.*, 1900, (4), iii. 159.
10. Schneebeli, *Wied. Ann.*, 1884, xxii. 430.
11. Wilson, *Astrophys. Jour.*, 1899, x. 80.
12. Valentiner, *Ann. der Phys.*, 1910, (4), xxxi. 272.
13. Lummer und Pringsheim, *Verh. der Deutsch. Phys. Gesell.*, 1899, i. 33 and 215.
14. Coblentz, Carnegie Inst. of Washington, 1906, Publ. No. 65, pp. 77 and 80, and 1908, Publ. No. 97, p. 18.
15. Coblentz, *Bur. Standards Bul.*, 1913, x. 7.
16. Waidner and Burgess, *Bur. Stds. Bul.*, 1907, iii. 163.
17. Coblentz, *Bur. Stds. Bul.*, 1908, vii. 169.
18. Coblentz, *Bur. Stds. Bul.*, 1912, ix. 106.
19. Lehnebach, *Pogg. Ann. der Phys.*, 1873, cli. 96.
20. Coblentz and Emerson, *Bur. Stds. Bul.*, 1916, xii. 503.
21. Christiansen, *Wied. Ann. der Phys.*, 1883, xix. 267.
22. Kurlbaum, *Wied. Ann.*, 1898, lxv. 746.
23. Kurlbaum, *Verh. d. D. Phys. Gesell.*, 1912, xiv. 576.
24. Coblentz, *Bur. Stds. Bul.*, 1913, ix. 283.
25. Valentiner, *Ann. der Phys.*, 1912, (4), xxxix. 489; 1913, xli. 1059.
26. Kurlbaum und Valentiner, *Ann. der Physik*, 1913, (4), xli. 1059.
27. Coblentz, *Bur. Stds. Bul.*, 1918, xiv. 532.
28. Gerlach, *Ann. der Phys.*, 1913, (4), xli. 99.
29. Coblentz, *Bur. Stds. Bul.*, 1916, xii. 553.
30. Callendar, *Proc. Phys. Soc.*, London, 1910, xxiii. 1.
31. Gerlach, *Ann. der Phys.*, 1912, (4), xxxviii. 1.
32. Ångström, *Phys. Rev.*, 1893, i. 365; *Ann. der Phys.*, 1899, (3), lxvii. 663.
33. Gerlach, *Ann. der Phys.*, 1916, (4), l. 233.
34. Paschen, *Ann. der Phys.*, 1912, (4), xxxviii. 30.
35. Gerlach, *Ann. der Phys.*, 1913, (4) ix. 701; xli. 99; xlii. 1163 and 1167.
36. Gerlach, *Ann. der Phys.*, 1916, (4), l. 245.
37. Gerlach, *Ann. der Phys.*, 1916, (4), l. 259.
38. Coblentz, *Bur. Stds. Bul.*, 1911, ix. 40.
39. Coblentz, *Bur. Stds. Bul.*, 1920, xv. 520.
40. Coblentz, *Bur. Stds. Bul.*, 1914, xi. 157.
41. Coblentz, *Bur. Stds. Bul.*, 1916, xiii. 423.
42. Christiansen, *Wied. Ann. der Phys.*, 1883, xix. 268.
43. Puccianti, *Nuovo Cimento*, 1912, (6), iv. 31.
44. Kahanowicz, *Nuovo Cimento*, 1917, (6), xiii. 142; *Accad. Lincei*, *Atti*, 1919, xxviii. 73.
45. Féry, *Compt. Rendus*, 1909, cxlviii. 915.
46. Féry et Drecq, *Jour. de Phys.*, 1911, (5), i. 551.
47. Féry et Drecq, *Compt. Rend.*, 1912, clv. 1230.
48. Bauer, *Soc. Franç. de Phys.*, Procès Verbaux, 1913, No. 41.
49. Bauer et Moulin, *Compt. Rend.*, 1909, cxlix. 988; *Bull. Séances Soc. Fr. de Phys.*, 1910, p. 58.
50. Bauer et Moulin, *Compt. Rend.*, 1910, cl. 167.
51. Puccianti, *Nuovo Cimento*, 1912, (6), iv. 322.
52. Keene, *Proc. Roy. Soc.*, 1913, lxxxviii. A, 49; *London Electrician*, 1912, lxx. 541.
53. Shakespear, *Proc. Roy. Soc.*, London, 1912, lxxxvi. A, 180.
54. Todd, *Proc. Roy. Soc.*, 1919, lxxxiii. A, 19.
55. Westphal, *Verh. Phys. Gesell.*, 1912, xiv. 987.
56. Westphal. *Verh. d. Deutsch. Phys. Gesell.*, 1913, xv. 897.
57. Lewis and Adams, *Phys. Rev.*, 1914, (2), iii. 92.
58. Millikan, *Phys. Rev.*, 1916, (2), vii. 379.
59. Paschen, *Sitzber. Akad. Wiss.*, Berlin, 1899, xxi. 403, 959.
60. Paschen, *Ann. der Phys.*, 1901, (4), iv. 277.
61. Lummer und Pringsheim, *Verh. Phys. Gesell.*, 1899, i. 23.
62. Lummer und Jahnke, *Ann. der Phys.*, 1900, (4), iii. 283.
63. Lummer und Pringsheim, *Verh. Phys. Gesell.*, 1899, i. 215.
64. Holborn und Valentiner, *Ann. der Phys.*, 1906, (4), xxii. 1.
65. Valentiner, *Ann. der Phys.*, 1912, (4), xxix. 489.
66. Coblentz, *Bur. Stds. Bul.*, 1916, xiii. 459.
67. Paschen, *Ann. der Phys.*, 1913, xli. 670.
68. Rubens and Kurlbaum, *Ann. der Phys.*, 1901, (4), iv. 649.
69. Nernst und Wulf, *Verh. Phys. Gesell.*, 1919, xxi. 294.
70. Warburg, Hupka, und Müller, *Zs. Instrk.*, 1912, xxxii. 134.
71. Dellinger, *Bur. Stds. Bul.*, 1916, xiii. 535.
72. Roeser, *Bur. Stds. Bul.*, 1917, xiv. 237.
73. Paschen, *Ann. der Phys.*, 1894, (3), liii. 301; 1901, (4), iv. 299.
74. Lummer und Pringsheim, *Verh. Phys. Ges.*, 1900, ii. 163.
75. Lummer und Pringsheim, *Ann. der Phys.*, 1900, (4), iii. 159.
76. Holborn und Day, *Wied. Ann. der Phys.*, 1899, lxviii. 817; *Ann. der Phys.*, 1900, (4), ii. 505.
77. Mendenhall and Forsythe, *Phys. Rev.*, 1914, (2), iv. 62.
78. Mendenhall, *Phys. Rev.*, 1917, (2), x. 515.
79. Hoffmann und Meissner, *Zs. Instrk.*, 1913, xxxiii. 157.
80. Warburg und Leithauser, *Sitzber. Akad. Wiss.*, Berlin, 1910, xliv. 225.
81. Warburg und Leithauser, *Zs. für Instrk.*, 1910, xxx. 118.
82. Warburg und Leithauser, *Zs. für Instrk.*, 1911, xxxi. 124.
83. Warburg, Hupka, und Müller, *Zs. für Instrk.*, 1912, xxxii. 134.
84. Warburg, Leithauser, Hupka, und Müller, *Sitzber. Akad. Wiss.*, Berlin, 1913, i. 35.
85. Warburg, Leithauser, Hupka, und Müller, *Ann. de Phys.*, 1913, (4), xl. 609.
86. Warburg, Hupka, und Müller, *Zs. Instrk.*, 1914, xxxiv. 115.
87. Warburg und Müller, *Zs. Instrk.*, 1915, xxxv. 98.
88. Warburg und Müller, *Ann. der Phys.*, 1915, (4), xlviii. 410.
89. Warburg, *Verh. Phys. Ges.*, 1916, xviii. 1.
90. Hoffmann und Meissner, *Ann. der Phys.*, 1919, (4), lx. 201.
91. Holst and Oosterhuis, *Proc. K. Akad.*, Amsterdam, 1917, xix. 549.
92. Hyde and Forsythe, *Astrophys. Jour.*, 1920, li. 244.
93. Carvallo, *Compt. Rend.*, 1898, cxxvi. 728.
94. Langley, *Annals Astrophys. Obs.*, Smithsonian Institution, 1902, i. 222, Plate XXVI. B.
95. Michell, *Ann. der Phys.*, 1902, (4), vii. 772; Liebreich, *Verh. Phys. Gesell.*, 1911, xiii. 1.
96. Paschen, *Ann. der Phys.*, 1908, (4), xxvi. 120.
97. Paschen, *Ann. der Phys.*, 1901, iv. 299; Pl. II.
98. Blake and Duane, *Phys. Rev.*, 1917, (2), x. 624.
99. Hennig, *Verh. Deutsch. Phys. Gesell.*, 1918, xx. 81.
100. Birge, *Phys. Rev.*, 1919, (2), xiv. 361.
101. Coblentz, *Bur. Stds. Bul.*, 1920, xv. 617.
102. Sommerfeld, *Atombau und Spektrallinien*, 1919.
103. Gerlach, *Zs. für Physik*, 1920, ii. 76.

RADIATION THEORY

§ (1).—The behaviour of electromagnetic radiations can be regarded from two very different points of view. In optics the interest attaches to the study of the light waves themselves—their refraction, interference etc.—and little is said of the way in which the light arises. The present problem is the complement of this. It is concerned almost entirely with the emission and absorption of light, and makes use of optical properties only in so far as they help in the study of these. In consequence of this difference of aspect the mode of treatment is quite different. Physical optics is based on dynamics (including electromagnetic theory under the term), but the theory of radiation is founded on thermodynamics. As will be seen, pure thermodynamics can give a great deal of information about radiation, but is not capable of completely solving the problems that arise, just as it gives a great deal of information about the behaviour of a gas, but is not competent to give completely its equation of state. The subsequent problem is therefore to be solved by other methods. A direct application of the principles of mechanics leads to a definite and complete answer, but one which is entirely wrong; and to overcome this difficulty Planck introduced the quantum[1] principle, with the help of which the central problem of the theory is solved.

§ (2) Definitions.—It will be necessary to make certain definitions for the measure of radiation. These naturally take a very different form from those which are used in optics. There the starting-point is usually a train of monochromatic, polarised, plane, or spherical waves, and consequently the fundamental quantity is the amplitude of the waves. But in thermodynamics this is of no interest. The fundamental quantity is heat or energy, and the radiation measured as energy must be supposed to have a continuous distribution of frequencies and to be travelling in all directions.

The suitable definition is this. Take a small aperture $d\sigma$ and draw from its edges a cone of small solid angle $d\omega$; let the axis of the cone be at angle θ to the normal of $d\sigma$. Then a certain part of the radiation will pass through this aperture into the cone in time dt. If, as we suppose, the radiation is continuously distributed, the amount must be proportional to $d\sigma \cos\theta d\omega dt$ and the total energy entering the cone as radiation may be taken as $K d\sigma \cos\theta d\omega dt$. But a further analysis is necessary, for the pencil of rays may be supposed analysed into a spectrum. This spectrum will be continuous and so the definition will be concerned with the amount of energy between frequencies ν and $\nu + d\nu$. Finally, the radiation may be partly polarised, and this is represented by choosing two fixed directions and describing the energy corresponding to waves polarised in each. Thus the description of the radiation is fully given by an expression of the form

$$(K_\nu + K_\nu')d\nu d\sigma \cos\theta d\omega dt;$$

if the components of polarisation K_ν, K_ν' are known at every point and time the whole radiation field is fully described.

The connection between this definition and those of optical theory is not obvious. The definition is of course quite unsuited for the type of waves usually considered in theoretical optics, as these have been simplified by being taken monochromatic, plane, etc. (which is only approximately true), and the result of the present definition would be that K_ν would be infinite for one exact set of values and zero for all others. But if the electric and magnetic vectors are arbitrarily given at every point in a space, it is possible to deduce K_ν from them by means of rather complicated applications of Fourier integrals. The chief point of interest in the work lies in the spectral resolution into frequencies, which is done by virtue of a theorem of Lord Rayleigh's, whereby from a Fourier analysis of *amplitudes* it is possible to define a precise meaning for the *energy* in any range of frequencies.

In the treatment of radiation problems by thermodynamics it will be necessary to have idealised machinery for sorting out the radiation into its component parts. For example, the polarisation may be studied by means of Nicol prisms, but these have to be supposed capable of completely transmitting one component and completely reflecting the other. More important is the question of the spectral analysis. For this, advantage may be taken of the fact that many substances show a selective effect, transmitting light of some frequencies and reflecting the rest. If this is idealised it may be imagined that there exists a set of screens, each of which has the property of transmitting a certain range of frequencies perfectly while reflecting all others perfectly. The justification for the use of this idealised machinery here is much the same as in other branches of thermodynamics, as, for example, the use of ideal semi-permeable membranes in the study of mixtures of gases.

Next, consider the emission of radiation. This is defined in the same sort of way as above. A small volume dv of a body emits in time dt radiant energy $(\epsilon_\nu + \epsilon_\nu')d\nu d\omega dv dt$ into a solid angle $d\omega$. In the general case of an anisotropic body the two components of polarisation ϵ_ν and ϵ_ν' will both be functions of the direction of the element of solid angle. The emission from any substance in general

[1] See "Quantum Theory."

will depend on its past history (as in phosphorescence) or on the forces acting on it (as in tribo-luminescence and the light from discharge tubes). Now such effects as these are not steady state phenomena and so must be excluded from thermodynamic arguments. We shall exclude all such effects by the assumption that ϵ_ν and ϵ_ν' depend only on the nature of the matter composing dv, and on the temperature. Radiation which satisfies these conditions is called *temperature radiation*. The assumption includes what is known as Prevost's law of exchanges, viz., that a body emits an amount of radiant energy that depends only on its own temperature and not on that of its surroundings.

For absorption the type of definition is rather different. The absorption coefficient a_ν for any matter is defined by the fact that if a beam traverses a short length dl of it, then the beam's intensity is reduced by a fraction $a_\nu dl$ of itself, and the part which has disappeared from the radiation reappears as heat in the matter. For an anisotropic body a_ν will depend on the direction and polarisation of the beam. As a corollary, if a ray traverses a length l of matter its emergent intensity is a fraction $e^{-a_\nu l}$ of its incident.

§ (3) Radiation in the Steady State.—With these definitions it is possible to deduce many important consequences simply from the fact that a system left to itself tends to equalise in temperature and so arrive at a steady state. The proofs of these propositions are quite elementary and it will be unnecessary to give the chain of reasoning by which they are obtained.

(i.) The K_ν of a pencil of rays is invariant along its path in free space.

(ii.) K_ν is invariant for total reflection at any surface, plane or curved. (The concentration of light by a concave mirror is not due to any change in K_ν, but to an increase in the total solid angle through which the rays arrive at the focus.)

(iii.) In a non-homogeneous medium let the waves of frequency ν have refractive index μ_ν, varying from point to point. Then K_ν/μ_ν^2 is invariant along a pencil of rays.

These three results are really only matters of geometry.

(iv.) At a surface separating two media partial reflection occurs. Take an incident ray along the direction A, and suppose that a fraction ρ is reflected and a fraction $1-\rho$ refracted along B in the second medium; then the possibility of a steady state requires that the reverse ray coming along B should have a fraction ρ reflected and a fraction $1-\rho$ refracted into the first medium along A reversed.

A specially important case of this is that an ideal selective screen, which transmits light of one set of frequencies and reflects all others, must do so in exactly the same way from whichever side the light is coming. The same line of argument shows that it is impossible to combine the light from two sources by any optical device, so as to increase the K_ν of a pencil of rays. The results of (iv.) can be proved also on dynamical principles; the advantage of the present method is that it makes no reference to the mechanism of reflection and refraction.

(v.) Consider a vacuous space surrounded by perfectly reflecting insulating walls. In it are certain bodies of any character. By means of their radiations to one another they gradually set themselves to the same temperature. A test body placed anywhere in the space must also reach this temperature, and that through the action of radiation alone. By taking the test body as covered in turn with various types of selective screens, Nicol prisms, etc., it can be shown that the value of K_ν is the same at every point of the space and for every direction and polarisation of the rays. (If the refractive index differs from unity, K_ν must be replaced by K_ν/μ_ν^2.) This is the principal result of the argument—that in the steady state the radiation K_ν can depend only on ν and the temperature, and must be quite independent of the position, etc., of the ray examined. Such radiation is called *complete radiation*; also, for a reason that will appear later, *black radiation*. The determination of the form of K_ν as a function of ν and T is the cardinal problem of the radiation theory.

It is often convenient, in the general case where the radiation is not *complete*, to assign a meaning to the expression "temperature of a ray." This is defined as the temperature of the complete radiation which would give the same K_ν as the ray. The temperature of a ray can never be greater than that of the source from which it arose; this assumes that the radiation is temperature radiation. Conversely a body can never get hotter than the temperature of the hottest rays which strike it. To make a body as hot as possible by radiation it would be necessary to arrange some optical device whereby the hottest rays should strike it equally from all directions, while colder rays, of other frequencies, should be excluded by selective screens.

As the complete radiation is isotropic, it is convenient to use a quantity which makes no reference to the directions in which the rays are going. This is done by replacing the quantity K_ν by a derived quantity u_ν, where $u_\nu d\nu$ is the total energy in unit volume of frequencies between ν and $\nu+d\nu$. By summing the effects of all the rays passing through a small element of volume, it is easy to show that $u_\nu = 8\pi K_\nu/c$, where c is the velocity of light. The factor is 8π instead of the more usual 4π, because there are two components

of polarisation each with K_ν. u_ν will be the same at all points of the enclosure. (For the case with a refractive index the corresponding equation is $u_\nu = 8\pi K_\nu \mu_\nu / c$, and u_ν/μ_ν^3 will be the invariant.)

§ (4) KIRCHHOFF'S LAW.—In consequence of the universality of K_ν there must be a certain relation between the emission and absorption of every type of matter. Take a small body of volume v. A ray K_ν, in traversing it, loses an amount of energy $a_\nu K_\nu d\nu d\omega dtv$, by the definitions of § (2), and as the radiation must be the same on both sides of the body, this must be replaced by emission. The amount of the emission is $\epsilon_\nu d\nu d\omega dtv$. Therefore $\epsilon_\nu = a_\nu K_\nu$, and either $K_\nu = \epsilon_\nu / a_\nu$ or else both ϵ_ν and a_ν are zero. This result, Kirchhoff's law, may be stated as follows. The ratio of the coefficient of emission to the coefficient of absorption is the same for all substances whatever, and depends only on the temperature and the frequency of the rays examined. As long as only temperature radiation is considered this law admits of no exceptions. It is proved in the first place for bodies in an enclosure; its great utility lies in its extension to free space by virtue of Prevost's law of exchanges.

Kirchhoff's law makes it possible to deduce certain interesting and not very obvious consequences. For example, tourmaline polarises light by absorbing one component strongly, while transmitting the other. To the high absorption must correspond a high emission; so if a tourmaline crystal is heated, as it preserves the same characteristic when hot, it will emit polarised light, and the plane of polarisation will be perpendicular to that of the light which the crystal ordinarily transmits. Again, nebulae of a size thousands of times as large as the solar system emit light confined to a few lines in the spectrum. Therefore their matter is so rare and of such a peculiar character that a ray of light of any other frequency can go right through this enormous distance without experiencing any absorption whatever.

Next, suppose that we have a substance with the property that all radiation falling on its surface is completely absorbed. This is called a *black body*. Perfect blackness requires that there shall be no reflection at the surface, and this implies no change of refractive index there, and also that a_ν, the absorption coefficient, should be infinitesimal; the latter condition requires that the body should be of infinite depth. Such a surface has the property of emitting the complete radiation, whether enclosed in an envelope or in free space. For consider the radiation going into a small cone $d\omega$ at angle θ to the normal. One of the polarised components emitted by an element of volume $d\sigma dz$ at depth z below the surface is $\epsilon_\nu d\nu d\omega dt d\sigma dz$, and of this only a fraction $e^{-a_\nu z \sec\theta}$ gets through the surface. The beam outside is obtained by integrating this over all depths, and so is $(\epsilon_\nu/a_\nu) d\nu d\omega dt \cos\theta d\sigma$. Hence by Kirchhoff's law the K_ν outside is exactly the same as what it would be in an enclosed space at the same temperature as the black body; this explains why *complete radiation* is often called *black*.

In consequence of the presence of the factor $\cos\theta$ in the above formula a black surface will appear equally bright viewed from any angle. Any departure from the condition of blackness destroys this. For example, a change of refractive index at the surface involves partial reflection, and it is well known from optical theory that the oblique rays are more reflected than the perpendicular. Therefore such a surface will emit less light obliquely than perpendicularly.

The importance of the idea of the black body is that it makes it possible to get away from the highly idealised arguments about radiation in a completely enclosed space, and so to connect pure theory with the results of observation. To find the experimental values for the complete radiation it will be sufficient to determine the energy emitted (in various frequencies) from a black body of known temperature. In order to find its amount this energy must be completely absorbed. Thus both source and receiver must be black. The best form of source is a small hole in the wall of a furnace; for the absorption of such a hole is nearly perfect, since any ray that enters will undergo many reflections before coming to the hole again, and so will be practically completely absorbed, and therefore the emission of such a hole must also be nearly that of a black body. The same form is sometimes used for the receiving instrument; but in cases where there is little intensity of radiation it is necessary to concentrate the heat absorbed on as small an area as possible, so as to get the greatest temperature rise. It is therefore covered with platinum black or lamp black, and by subsidiary experiments an estimate is made of the imperfection of its blackness. For this part of the subject see "Radiation."

§ (5) STEFAN'S LAW AND WIEN'S LAW.—The theorems hitherto obtained were all based simply on the idea of equalisation of temperature. By the use of the second law of thermodynamics in the form that perpetual motion is impossible, two very important theorems can be proved—the law of Stefan and Boltzmann, and the displacement law of Wien. These are both concerned with complete radiation, and it will be convenient to replace the fundamental K_ν by the derived quantity u_ν, which depends on the energy per unit volume.

(i.) The total radiation of all frequencies in unit volume is denoted by $u=\int_0^\infty u_\nu d\nu$. It is a function of the absolute temperature T only and the Stefan-Boltzmann law [1] asserts that it is proportional to the fourth power of T.

First observe that the second law of thermodynamics requires that radiation should exert a pressure. Otherwise it would be possible to concentrate the radiation inside a perfectly reflecting cylinder by pushing in the piston, without doing any work; and the concentrated energy could be absorbed by a body hotter than that from which it was emitted. From the electromagnetic theory Maxwell showed that this pressure is equal to $\frac{1}{3}u$ for isotropic radiation and this leads to the result sought.

Consider a cylinder of volume V with perfectly reflecting walls. At one end in it is placed a small piece of matter to act as a vehicle for the transfer of heat from outside into radiant energy in the cylinder. The total energy of radiation is Vu. Now suppose that the volume is increased by an amount dV, while at the same time heat flows in from the outside and is partly converted into radiation by the piece of matter. The amount of heat absorbed will be

$$\mathrm{C}d\mathrm{T}+d(\mathrm{V}u)+\tfrac{1}{3}ud\mathrm{V},$$

or

$$\left(\mathrm{C}+\mathrm{V}\frac{\partial u}{\partial \mathrm{T}}\right)d\mathrm{T}+\tfrac{4}{3}ud\mathrm{V},$$

where C is the heat capacity of the matter. Now following the ordinary thermodynamic argument, if this is divided by T the result must be a perfect differential, and for this to be so

$$\frac{\partial}{\partial \mathrm{V}}\left(\frac{\mathrm{C}}{\mathrm{T}}+\frac{\mathrm{V}}{\mathrm{T}}\frac{\partial u}{\partial \mathrm{T}}\right)=\frac{\partial}{\partial \mathrm{T}}\left(\frac{4}{3}\frac{u}{\mathrm{T}}\right),$$

which leads to $\partial u/\partial \mathrm{T}=4(u/\mathrm{T})$. If this equation is integrated the result is $u=\sigma \mathrm{T}^4$, where σ is a constant. This is Stefan's law. A further consequence which will be required later is that in an adiabatic expansion VT^3 is constant.

The argument can be reversed so that the law of pressure may be deduced from Stefan's law. Thus, if the pressure is supposed to be the unknown f, an equation of the form $(\partial f/\partial \mathrm{T})-f/\mathrm{T}=u/\mathrm{T}$ must hold. Now if $u=\sigma \mathrm{T}^4$ is substituted and the equation integrated the result is $f=\frac{1}{3}u+a\mathrm{T}$, where a is an integration constant. So apart from this constant Stefan's law leads straight to Maxwell's law of radiation pressure. It is far the most accurate verification of the validity of that law, as the pressure itself is so exceedingly minute as to be barely perceptible to direct observation.

(ii.) The relation of Stefan and Boltzmann concerns only the total radiation, and has nothing to say as to the distribution in the various parts of the spectrum. If thermodynamics are to be made to give any information about this, it is necessary to have some mechanism whereby radiation of one frequency can be changed reversibly into another. The usual mechanism of conversion is by absorption and re-emission, but as long as the relation $\mathrm{K}_\nu=\epsilon_\nu/a_\nu$ is satisfied the thermodynamic conditions are fulfilled and so this method gives no help. But there exists another way of changing frequencies, namely, by the Doppler effect; that is, by reflection of the rays at a moving mirror. By the use of this idea it is possible to deduce a relation connecting u_ν, ν, and T. The argument falls into two parts—a thermodynamic and a mechanical. The former shows that, under certain conditions, if a reflecting enclosure is filled with *complete* radiation and then expanded adiabatically, the radiation will remain *complete*, without the intervention of matter. The second part consists in calculating the actual change in the radiation produced by the expansion.

(*a*) Consider the adiabatic compression of a cylinder with perfectly reflecting walls, supposing that some part of the walls reflects diffusely, so that the radiation certainly remains isotropic. At every stage the work done depends on the pressure, and this is given by the density of the total radiation. So whether the radiation remains *complete* or not, the energy density will be the same at every stage, and so will the work of compression. The relation VT^3=constant implies that when the volume is V the density of total radiation is the same as that of complete radiation of temperature T.

Now imagine that the effect of the adiabatic compression is to make the radiation no longer complete. Take V_1 as the initial volume, and suppose it filled with complete radiation of temperature T_1. Compress the cylinder slowly by a finite amount to volume V_2. Then if $\mathrm{V}_2\mathrm{T}_2^3=\mathrm{V}_1\mathrm{T}_1^3$, the density of total radiation will be the same as that of complete radiation of temperature T_2, but by hypothesis some spectral regions will be in excess and some in defect. Now introduce a small piece of matter of negligible heat capacity. This will readjust the radiation to completeness, but will not alter its total amount. The adjustment is an irreversible process and so involves an increase of entropy. Next, keeping the matter in the cylinder, slowly expand to the original volume. The work gained in the up-stroke is exactly equal to that done in the down-stroke, for they both depend only on the total radiation. If the matter is now withdrawn, the system will have returned to its original state, and no heat has been communicated from outside. The cycle could be repeated indefinitely and it would be possible to get a continual increase of entropy without any introduction of heat or performance of work; and this is contrary to thermodynamic principles. Therefore the adiabatic compression must conserve the completeness of radiation.

There is one small point in the proof worth mentioning. Reflection at a moving mirror not only changes the frequency, but also the direction of the light. If the walls are perfect cylinders with perpendicular ends, the compression will make the radiation anisotropic, and so of course not complete. This is the reason why it was necessary to specify that some part of the walls must reflect diffusely.

(*b*) It is now necessary to examine the effect on radiation when it is reflected at a moving mirror;

[1] See "Radiation," Part II.

in particular to count up the losses and gains for a small region $d\nu$ of the spectrum. If a beam inclined at angle θ falls on a mirror moving with small velocity v, it may be shown that the frequency of the reflected beam is $\nu'=\nu(1-2(v/c)\cos\theta)$ and that its direction is given by $\theta'=\theta+2(v/c)\sin\theta$. There is also a change in the intensity, since work is done on the mirror. It may be shown that the ray $u_\nu d\nu\,(\sin\theta d\theta d\phi/4\pi)$ (ϕ is the azimuthal angle) exerts a pressure

$$2\cos^2\theta\, u_\nu d\nu\,(\sin\theta d\theta d\phi/4\pi).$$

Now consider the gains and losses to the region $d\nu$. Of energy originally in this region and at this angle an amount

$$u_\nu d\nu\,(\sin\theta d\theta d\phi/4\pi)\,(c\cos\theta-v)dtd\sigma$$

strikes the mirror in time dt and in the element of area $d\sigma$. This energy is all lost. Summing over all angles the gross loss is

$$u_\nu d\nu\,\{(c/4)-(v/2)\}dtd\sigma.$$

There is a compensating gain from the light which has its frequency brought into $d\nu$. This comes from various regions. If one of them had frequency between ν_0 and $\nu_0+d\nu_0$ before reflection and inclination between θ_0 and $\theta_0+d\theta_0$, then $\nu_0=\nu\{1+2(v/c)\cos\theta\}$, and $\theta_0=\theta-2(v/c)\sin\theta$. The energy in this region gives after reflection an amount in $d\nu$ equal to

$$u_{\nu_0}d\nu_0\left(\frac{\sin\theta_0 d\theta_0 d\phi}{4\pi}\right)(c\cos\theta_0-v)dtd\sigma$$
$$-2\cos^2\theta u_\nu d\nu\left(\frac{\sin\theta d\theta d\phi}{4\pi}\right)vdtd\sigma,$$

the subtraction of the second term being due to the disappearance of energy in performing work. If $u_\nu+(\nu_0-\nu)(\partial u_\nu/\partial\nu)$ is written for u_{ν_0} and if the values of ν_0 and θ_0 are substituted and the result integrated over all angles, it gives as the gross gain

$$\left[u_\nu\left(\frac{c}{4}-\frac{v}{2}\right)+\frac{v}{3}\nu\frac{\partial u_\nu}{\partial\nu}\right]d\nu dtd\sigma.$$

The net gain is therefore

$$\tfrac{1}{3}\nu\frac{\partial u_\nu}{\partial\nu}d\nu vdtd\sigma,$$

and if this is summed over the whole mirror it becomes

$$\tfrac{1}{3}\nu\frac{\partial u_\nu}{\partial\nu}d\nu dV,$$

where dV is the total gain of volume in the time dt. Thus adiabatic expansion requires the equation

$$d(Vu_\nu)=\tfrac{1}{3}\nu\frac{\partial u_\nu}{\partial\nu}dV,$$

or written as a partial differential equation

$$\tfrac{1}{3}\nu\frac{\partial u_\nu}{\partial\nu}-V\frac{\partial u_\nu}{\partial V}=u_\nu.$$

The general solution of this equation is

$$u_\nu=\nu^3F(V\nu^3),$$

where F is an arbitrary function. But, since VT^3 remains constant in adiabatic expansion, $V=\text{constant}/T^3$, and this equation can be rewritten

$$u_\nu=\nu^3 f_1\left(\frac{\nu}{T}\right),$$

and this is Wien's displacement law.

The relation $u_\nu=\nu^3 f_1(\nu/T)$ makes it possible to deduce the radiation curve (that is, the curve connecting u_ν and ν) for any temperature once it is known for a single temperature. All purely thermodynamic conditions are now satisfied and the determination of the form of the function f_1 must invoke other principles. It is not surprising that it should not be possible completely to determine it; it is like the fact that no thermodynamic argument can completely determine the equation of state of a gas.

In connection with experimental work it is customary to use wave-length instead of frequency. This alters the form of the displacement law. Let

$$E_\lambda d\lambda=u_\nu d\nu,$$

so that E_λ is the density of isotropic energy per unit wave-length. If use is made of the relation $\lambda=c/\nu$ the law is given in the form

$$E_\lambda=\lambda^{-5}f(\lambda T).$$

§ (6) Planck's Radiation Formula.—The true radiation formula was found by Planck. Its theoretical deduction is given under "Quantum Theory."[1] The formula is

$$E_\lambda=\frac{8\pi hc}{\lambda^5(e^{hc/k T\lambda}-1)}.$$

Here c is the velocity of light—$2\cdot9986\times10^{10}$ cm. per sec.—and k and h are universal constants. k is the atomic gas constant; that is, $\frac{3}{2}kT$ is the mean kinetic energy of a molecule of a monatomic gas (supposed a perfect gas) at absolute temperature T. Its value is approximately $1\cdot372\times10^{-16}$ erg deg.$^{-1}$. h is the quantum, and its value is approximately $6\cdot55\times10^{-27}$ erg sec. Observe that the expression satisfies Wien's displacement law.

In *Fig.* 1 the firm line shows the curve for $T=1000°$ abs. The main characteristic of the curve is the rapid fall from the maximum on the side of short wave-lengths and the much more gradual fall on the other side. The unit for λ is 10^{-4} cm.—the visible spectrum would come about in the region between 0·4 and 0·8 of λ. E_λ is in C.G.S. units. Its significance will be made clearer by the statement that the energy in 1 c.c. in a spectral region of breadth 1 Å.U. (10^{-8} cm.) is $17\cdot38\times10^{-8}$ erg at the wave-length 28,850 Å.U. (the values are those at the maximum point).

The corresponding curve at a higher temperature T′ is obtained by shortening each λ in the ratio T/T′ and lengthening the corresponding E_λ in the ratio $(T'/T)^5$. The chain curve in *Fig.* 1 is that for the absolute temperature 1100°, and it may be seen what a large increase in the radiation is made by a

[1] See "Quantum Theory," § (4).

comparatively small change of temperature. Of two curves the hotter always lies entirely outside the colder.

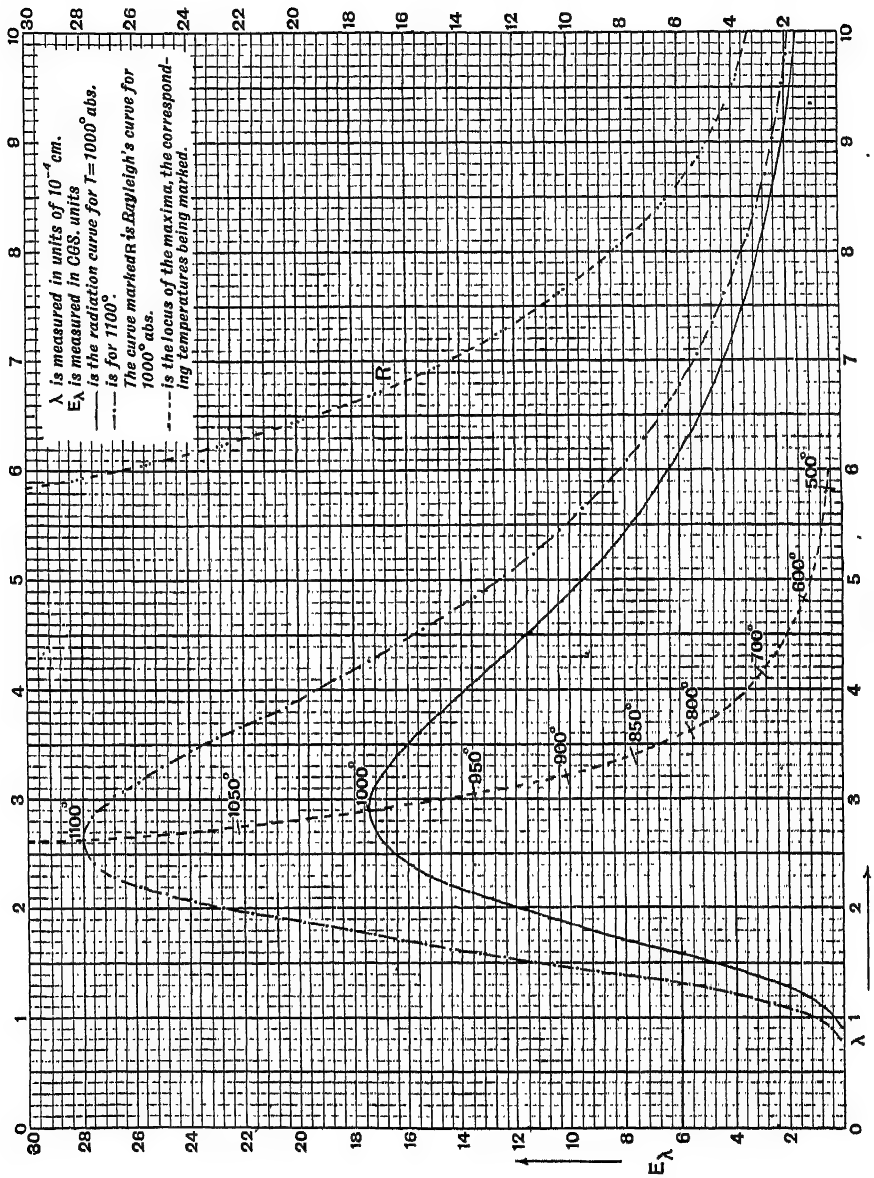


FIG. 1.

On the side of the short wave-lengths the curve is closely approximate to

$$E_\lambda = \frac{8\pi hc}{\lambda^5} e^{-hc/k T\lambda},$$

which is known as Wien's formula. The curve for this expression would lie very close to the corresponding curve of *Fig.* 1, except on the extreme right, where it would fall somewhat below it. Apart from the fact that it is a convenient approximation, the interest of Wien's expression is mainly historical, as it was the earliest radiation formula to be suggested.

On the side of the long wave-lengths, but

far beyond the region shown in the figure, Planck's formula becomes approximately

$$E_\lambda = \frac{8\pi k T}{\lambda^4}.$$

This is known as Rayleigh's formula.[1] The curve in *Fig.* 1 marked R is the Rayleigh curve corresponding to 1000°. Its importance lies in the fact that it is undoubtedly the expression which ought to be found if the classical system of mechanics were valid. For its derivation see "Quantum Theory," § (4). Planck found his formula by combining those of Wien and Rayleigh, so that each should be verified in its region of validity, and the formula then led him to his great discovery of the quantum.

The position of the maximum ordinate is given by

$$\lambda_m = \frac{hc}{kT}\frac{1}{\alpha},$$

where α is the solution of the equation

$$1 - e^{-\alpha} = \frac{\alpha}{5};$$

that is, $\alpha = 4{\cdot}9651.$

The height of the maximum is proportional to T^5. Its locus is shown in the broken curve of *Fig.* 1. It is of interest to observe the position of the maximum for a few temperatures. At ordinary room temperature it is about 10^{-3} cm. At 1000° abs., a dull red heat, it is still at a wave-length more than four times that of visible light. Even at 2000°, which is a dazzling white heat, it is still more than twice the wave-length of visible light. Again, from the fact that the sun's radiation has its maximum in the yellow, it may be deduced that its temperature is nearly 6000° abs.—a result confirmed by the application of Stefan's law to its total radiation.

Stefan's constant[2] σ may be deduced from Planck's formula by integrating over all values of λ. This leads to the result

$$\sigma = \frac{8\pi^5}{15}\frac{k^4}{h^3c^3} \text{ erg cm.}^{-3} \text{ deg.}^{-4}.$$

The determination of the constants k and h is usually done in the following two stages. One relation is found by measuring the total radiation. This gives σ. The second is got from the shape of the curve or the position of the maximum, which it may be seen determines hc/k. The universal constants k and h found in this way by the most refined experiments are given with about the same order of accuracy as by other methods.

Radiation theory is dealt with very clearly and thoroughly in Planck's *Die Theorie der Wärmestrahlung*, 4th edition (Barth, Leipzig). Many text-books of modern physics devote a chapter to the subject; for instance, Richardson's *The Electron Theory of Matter* (Camb. Univ. Press).

C. G. D.

[1] "Remarks upon the Law of Complete Radiation," *Phil. Mag.*, 1900, xlix. 539.
[2] See "Radiation," Part II.

Radiations from Radioactive Matter, effects of, on glass. See "Radioactivity," §§ (10), (15).
Various effects produced by. See *ibid.* § (15).

Radioactive Constant of a Radio-element: a term used in radioactivity to denote the fraction of the total amount of radioactive material changing in a unit of time, the unit of time being so chosen that the quantity of radioactive material at the end of it is sensibly the same as that at the beginning. See "Radioactivity," § (5).

Radioactive Substances, rays from, capable of exciting certain materials to emit visible light. See "Luminous Compounds," § (1).

RADIOACTIVITY

§ (1) General Properties of Radioactive Bodies.—The property of radioactivity was discovered by Henri Becquerel in 1896 for compounds of uranium, which he found to be spontaneously emitting radiations capable of affecting a photographic plate and of penetrating considerable thicknesses of matter. The radiations were also found to cause certain salts to fluoresce and to ionise air and other gases through which they passed. A fuller examination of the uranium salts showed that no variation could be detected in the intensity or in the character of the radiation with lapse of time. The intensity depended simply on the amount of uranium present, and was independent of the physical conditions to which the uranium was subjected. In 1898 Madame Curie and Schmidt independently showed that the property of radioactivity was also shared by thorium and minerals containing that element. The radiations from uranium and thorium are complex in character and consist of three distinct types which will be considered in detail later. These are known as the α, β, and γ radiations; the α rays consist of a stream of positively charged particles projected with great velocity, and are very easily absorbed by thin sheets of metal foil and by gases; the β rays are far more penetrating and are identical with the negatively charged particles constituting the cathode rays in a discharge tube; the γ rays are exceedingly penetrating and are identical in character with X-rays, the only difference between them being that the wave-length of the former is much shorter than that of the latter.

A substance which is capable of *spontane-*

ously emitting these penetrating radiations is said to be "radioactive." Of the 80 or more elements known in 1896 only uranium and thorium, the two heaviest elements, were found to be radioactive, and all the other radio-elements known to-day are derived from these two elements.

In addition to these radiations some of the radio-elements produce what are known as radioactive emanations which are gaseous. The emanations can be condensed at the temperature of liquid air, but even at ordinary temperatures they impart radioactivity to solid objects with which they come into contact.

In 1899 and 1900 the results of several investigations showed that some radioactive substances, unlike uranium, which showed no appreciable change of activity over a period of years, lost the greater part of their activity in the course of a few minutes or hours. For example, it was found that the emanation from thorium lost half its activity in less than one minute. Hence, in addition to the more permanent radioactive elements, there were others with only a transient existence.

§ (2) Uranium X.—Madame Curie had shown in her early experiments that the radioactivity of uranium was an atomic phenomenon. It is unaffected by chemical combination with other inactive elements. Crookes,[1] however, showed that by a single chemical operation, namely precipitating a solution of uranium with ammonium carbonate, the uranium could be obtained photographically inactive while the whole of the activity could be concentrated in a small residue free from uranium. This residue, which was called Uranium X, was many hundred times more active photographically weight for weight than the uranium from which it had been separated. A similar result was obtained by Becquerel, who found that barium could be made photographically very active by adding barium chloride to a uranium solution and precipitating the barium as sulphate. After a number of precipitations the uranium was rendered photographically inactive while the barium was strongly active. These results seemed to point to the fact that the activity of uranium was not due to the element itself but to some other substance which was associated with it.

The active barium and the inactive uranium in the last experiment were left for a year and again examined. It was now found that the uranium had completely regained its activity, while the activity of the barium sulphate had completely disappeared. The loss of activity of uranium was therefore only temporary in character.

§ (3) Thorium X.—Rutherford and Soddy[2] carried out similar experiments on thorium and were able in a single chemical operation to separate an intensely active constituent from thorium. This they called Thorium X. In a month's time the thorium X had lost all its activity whilst the thorium had completely regained it.

§ (4) Decay and Recovery of Uranium X.—The next step was to examine the time rate at which the processes of decay and recovery of activity took place. Uranium X emits only β rays, whilst uranium gives out only α rays. If all the measurements are made with β rays, the ionisation produced will depend on the quantity of uranium X present and not on the quantity of uranium—the uranium will be effective only in so far as it produces uranium X.

The uranium is left for some months so as to come into equilibrium with its product uranium X; they are then separated by one of the methods already mentioned. The β ray activity of the uranium will at first be zero, but it will gradually increase as uranium X is formed, whilst the activity of the uranium X falls off according to an exponential law. The two curves shown in *Fig.* 1 show the rates at

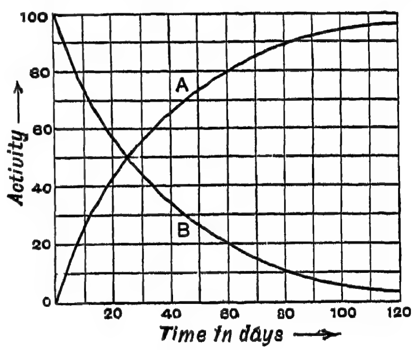


Fig. 1.

which the decay (curve B) and recovery (curve A) of uranium X take place. The sum of the β ray activities of the uranium and the uranium X is constant, showing that uranium X is being produced from uranium at exactly the same rate as it is decaying when separated from uranium. The curves show also that it decays to half value in 25 days.

Similar results have been obtained with other radioactive products separated from their parent elements, and the law that the rate of decay of activity of a given product is exactly equal to the rate of recovery of activity of the substance from which it has been separated, is universally true.

The decay curve in *Fig.* 1 shows that uranium X loses its activity according to an exponential law with the time. It will be seen later that

[1] Crookes, *Roy. Soc. Proc.*, 1900, A, lxvi. 409.
[2] Rutherford and Soddy, *Phil. Mag.*, 1902, iv. 370 and 569.

this is the general law of decay of activity in any type of radioactive matter, separated from its parent and from any secondary active products which it may itself produce. When this law is not fulfilled it can be shown that the activity is due to the superposition of two or more effects, each of which decays according to an exponential law with the time.

§ (5) THE RADIOACTIVE CONSTANT.—If N_0 is the number of atoms of any given product present at any time taken as zero, and N_t the number remaining after an interval of time t, then according to the above law of decay

$$N_t = N_0 e^{-\lambda t}, \quad . \quad . \quad . \quad (1)$$

where λ is a definite constant characteristic of the given product. The above equation may be written,

$$\frac{dN_t}{dt} = -\lambda N_t, \quad . \quad . \quad . \quad (2)$$

so that λ represents the fraction of the total amount of radioactive material changing in a unit of time, the unit of time being so chosen that the quantity of radioactive material at the end of it is sensibly the same as that at the beginning. It has different values for different types of radioactive material, but is invariable for any particular type of material. This constant λ is called the *radioactive* or *transformation constant* of the radio-element.

§ (6) THE HALF-VALUE PERIOD.—There is another constant which is extensively used in radioactivity. This is the *half-value period* or the *period of half change* of a radioactive substance. This is the time taken for the atoms present in a radio-element to decrease to half their number. Denoting this by T, equation (2) gives

$$T = 0 \cdot 6932 \, . \, \frac{1}{\lambda}. \quad . \quad . \quad . \quad (3)$$

The decay curve of uranium X given in *Fig.* 1 will be represented by equation (1) above, if the correct value is substituted for the radioactive constant λ. The equation for the recovery curve is at once obtained from the fact that at any time the sum of the quantity of uranium X in the portion separated from the uranium and the quantity produced by the uranium is a constant. It is

$$N_t = N_0(1 - e^{-\lambda t}), \quad . \quad . \quad . \quad (4)$$

where N_0 is the number of atoms of uranium X finally produced by the uranium.

The rate of decay of a radio-element is absolutely independent of any variation in physical and chemical conditions. For example, the decay of activity of any product takes place at the same rate when exposed to light as when it is kept in the dark, and at the same rate in a vacuum as in air or any other gas at atmospheric pressure. The activity is unaffected by intense heat or extreme
Electric discharges and strong magnetic f
are quite ineffective in producing a chan
the rate of decay.

§ (7) DISINTEGRATION THEORY.—The
duct uranium X is one example of radioa
matter of which there are many other t
Each such product has definite chemical as
as radioactive properties which distingui
not only from the other active products,
also from the substance from which
produced. To explain the continuous
duction of radioactive matter Rutherford
Soddy[1] in 1903 put forward the view tha
atoms of the radio-elements are underg
spontaneous disintegration, and that
disintegrated atom passes through a succe
of well-marked changes, accompanied by
emission of characteristic radiations.
theory has been found to account in a s
factory way for all known facts of radioacti

The general law governing the rate of d
or, in other words, the rate of disintegrati
all radioactive substances (given by equ
(1)) states that the number of atoms brea
up per unit time is proportional to the nu
of atoms present. The number of a
breaking up in a given time is subjec
fluctuations round the average value of
magnitude to be expected from the ge
probability theory, so that λ, the radioa
constant, represents the *average* fraction o
number of atoms which break up per
time. The fraction of a product transfo
per second is independent of the age of
product and does not depend upon the
centration of the active matter itself.
found that radium emanation, for insta
more than three months old decays at ex
the same rate as emanation freshly prod
by radium; the chance of an atom brea
up in a given time is the same whether
produced a second before or has existed i
pendently for a long period of time. T
facts lead to the conclusion that radioacti
as far as the individual atom is concerne
an instantaneous phenomenon. There i
gradual loss of energy. Before and up to
actual moment of disintegration the atom
radio-element is in no way different fro
atom of an inactive element. Similarly
new atom produced after disintegratio
similar in all respects to an ordinary a
Hence any atom may exist unchanged fo
time from zero to infinity.

§ (8) THE PERIOD OF AVERAGE LIFE O
ATOM.—It is often convenient to speak o
average life of a large number of atoms.
be the number of atoms present at the t
then after an interval of time t the nu
which change in the time δt is equal to
(from equation (2) above) or $\lambda N_0 e^{-\lambda t} \delta t$.

[1] Rutherford and Soddy, *Phil. Mag.*, 1903, v.

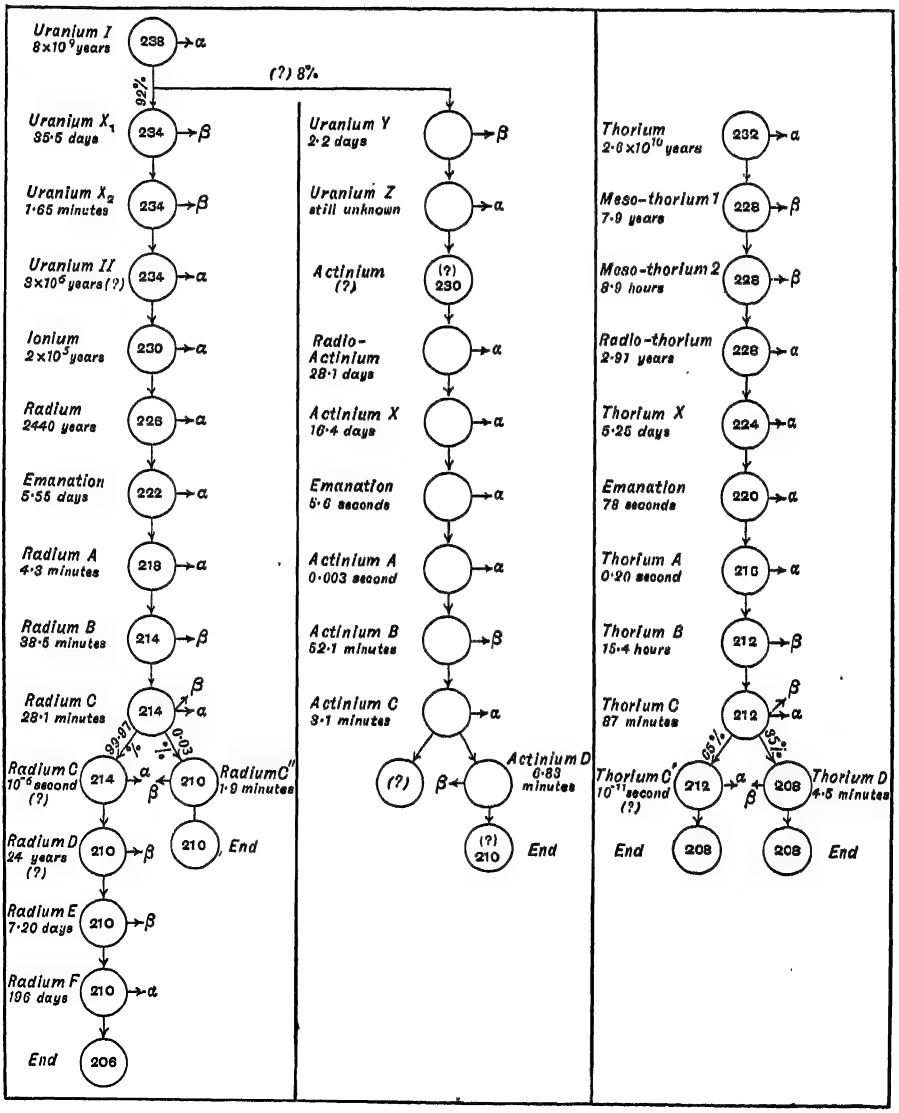
of these atoms has a life t, so that the average life of the total number is given by the expression

$$\int_0^\infty \lambda t e^{-\lambda t} \, . \, dt,$$

and this is equal to $1/\lambda$, that is, the period of average life of an atom is the reciprocal of the radioactive constant. The *period of average life of an atom* based on the above calculation may be defined as the sum of the separate periods of future existence of all the individual atoms divided by the number in existence at the starting-point, any time being taken as the starting-point. The constant relates, therefore, to the future life of the atom and is independent of the period the atom has already been in existence.

The period of average life of each of the atoms in the three disintegration series—radium, actinium, and thorium—is included in the following table, which also gives the atomic weight of the product together with the radiation emitted by it.

TABLE I



§ (9) Radioactive Equilibrium.—Each of the three disintegration series given in Table I. consists in a long succession of changes, P producing Q, Q producing R, and so on. The rate of change varies from product to product, but in each case it follows an exponential law. If a radioactive mineral is sealed up so that the products of transformation are allowed to accumulate, a stage is ultimately reached in which the amount of each product formed is equal to the amount transformed per unit time throughout the series. In this state there is a balance between the loss of activity of the matter already produced and the gain of activity due to the production of new active matter. Material which has reached this state is said to be in *radioactive equilibrium.*

It is evident that in the state of radioactive equilibrium the number of atoms disintegrating per unit time is the same for each product. So that if P, Q, R are the amounts of the successive members of the series present when radioactive equilibrium obtains, and λ_1, λ_2 and λ_3 their respective transformation constants, then

$$\lambda_1 P = \lambda_2 Q = \lambda_3 R. \quad . \quad . \quad . \quad (5)$$

The amount of each product present is therefore inversely proportional to its radioactive constant or directly proportional to its period of average life. According to the above relation there is a constant proportion between the quantities of successive members of a system in radioactive equilibrium. Thus in old uranium minerals the ratio of the amount of radium to uranium is always the same, there being about 3·2 tons of uranium for every gramme of radium. Now the period of average life of radium is known to be 2440 years, so that by the above relation the period of average life of uranium must be about 8×10^9 years. This relation may in this way be used to calculate the period of a long-lived product when its amount, relative to another whose period can be measured, is known. It is useful also to measure the quantity of a short-lived product when its period can be determined. Hence if the period and amount of a product in a given series are known and the period or amount of another product in the series can be measured, the remaining unknown quantity for this product can be calculated. In the radium series both the period and amount of emanation can be measured accurately with the result that the same information can be obtained by calculation for most of the other members of the series and for uranium itself.

The activity of a product is measured by means of effects produced by the radiations which it emits, and before proceeding further it would be advisable to consider the properties of these radiations in detail.

§ (10) Radiations of Radioactive Substances.—The first analysis of the complex radiations emitted by the radio-elements was made by Rutherford.[1] Two general methods were employed to distinguish between the types of radiations given out by the same body and to compare the radiations from different radio-elements. These were based on:

(i.) The deflection of the rays in a magnetic or in an electric field.

(ii.) The relative absorption of the rays in solids and gases.

Experiments carried out on these lines revealed three distinct types of radiation emitted by the radio-elements, which Rutherford called the α, β, and γ rays respectively. The difference between these radiations is brought out very clearly when a radioactive substance emitting the three kinds is placed in a strong magnetic field. Suppose a small quantity of radium in equilibrium with its products is placed at the bottom of the central hole in a lead cylinder L which rests on a photographic plate P. If a strong magnetic field be applied at right angles to the plane of the paper and directed towards the paper the three types are separated from one another in the manner shown diagrammatically in the figure. The γ rays proceed without deviation.

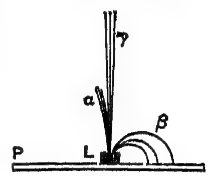


Fig. 2.

The α rays are deflected slightly to the left whilst the β rays move to the right, describing circular orbits whose radii vary within wide limits. The β rays are so much bent round that they strike the photographic plate, producing a diffuse impression upon it. The same relative effects are produced by the application of an electric field. It should be mentioned, however, that the deviation produced in the α ray beam is exceedingly small compared with that produced in the β ray beam. So small is the deviation produced in the case of the α rays that for some time they were thought to be non-deviable by a magnetic field until Rutherford[2] in 1903, by using a more active preparation of radium than had been previously employed, showed that they were deflected both by a magnetic and by an electric field. These experiments show that the α particles carry a charge opposite to that carried by the β particles. The latter were found by Giesel to be similar to the cathode particles in a discharge tube and therefore carried negative charges, whilst the former are similar to the canal rays in a discharge tube which had been shown by Wien to consist of positively charged particles travelling at great velocities.

[1] Rutherford, *Phil. Mag.*, 1899, xlvii. 116.
[2] *Ibid.*, 1903, v. 177.

§ (11) α Rays. (i.) *Range of α Rays.*—The α rays produce intense ionisation along their path in a gas, and, in consequence, they rapidly lose their kinetic energy until their velocity is reduced below the value at which they can ionise. Bragg and Kleeman [1] showed that the ionisation due to a homogeneous pencil of α rays ends after the rays have traversed a certain distance in air, this distance being called the range of the α particle in air. The range of an α particle from a simple product is a constant for a definite temperature and pressure of the gas traversed. It varies inversely as the pressure and directly as the absolute temperature, so that in specifying the range it is important to state the temperature and pressure as well as the nature of the gas. Different products emit α rays of different ranges, so that the range of the α particle is characteristic of the product from which it is emitted.

Geiger and Nuttall [2] employed the following method for determining the range, and it is applicable to every kind of radioactive matter provided the latter is not gaseous. The active material is placed in the form of a thin film on a small metal disc D in the centre of a metal bulb whose internal diameter is about 8 cm.

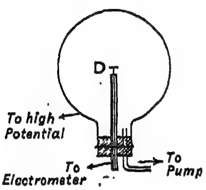


Fig. 3.

The disc, which is insulated from the bulb, is connected to one pair of quadrants of an electrometer, and the metal bulb is connected to a potential sufficiently high to produce the saturation current. Through the insulated stopper also passes a tube by means of which the bulb can be exhausted. The method of procedure was to measure the saturation current at different pressures. The results obtained are shown in *Fig.* 4. For low pressures the ionisation is very nearly proportional to the pressure, but when the pressure has reached a value such that all the α particles are completely absorbed in the gas, the ionisation current reaches a maximum value which remains constant with further increase of pressure, except in the case when two products in equilibrium were examined, in which case there are two abrupt breaks in the curve. The pressures at which the breaks in the curves occur correspond to the maximum ranges of the α particles in the gas at those pressures. The knowledge of this pressure enables the range in air at atmospheric pressure to be deduced, since the range is inversely proportional to the pressure.

[1] Bragg and Kleeman, *Phil. Mag.*, 1905, x. 318.
[2] Geiger and Nuttall, *ibid.*, 1911, xxii. 613; 1912, xxiii. 439; xxiv. 647.

Geiger [3] had previously found that the following relation existed between the range R of the α particle and its velocity V—

$$R = aV^3, \quad . \quad . \quad . \quad . \quad (6)$$

i.e. the range is proportional to the cube of the velocity. The results of Geiger and Nuttall showed that a definite relation existed also

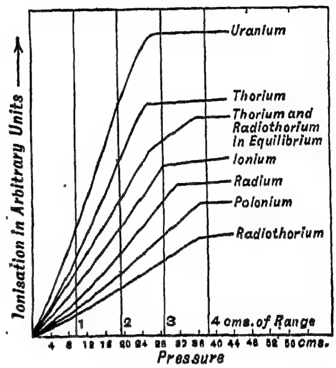


Fig. 4.

between the range and the radioactive constant of the substance emitting the radiation. This is shown graphically in *Fig.* 5.

If the range of the α particle of any product is known the period of this product can be deduced from this relationship. Thus the period of average life of uranium II should be about 3×10^6 years, and that of ionium 3×10^5

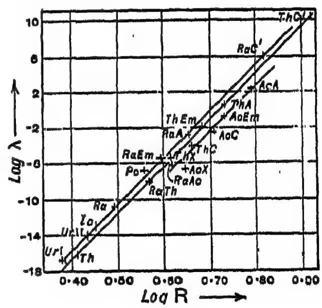


Fig. 5.

years, periods too long for direct determination. Similarly from the long range α particles emitted by radium C′ and thorium C′, the very short periods of these products can be deduced; these would be of the order of 10^{-6} and 10^{-10} seconds respectively.

The curves show that those products whose average life is long emit α rays whose range is small, and *vice versa*.

[3] Geiger, *Roy. Soc. Proc.*, 1910, A, lxxxiii. 5

(ii.) *The Stopping Power of Metal Foils.*—If a thin metal foil is inserted in the path of an α particle, the distance it will travel in air will be reduced. The difference between the range of the particle with and without the foil interposed is called the *stopping power* of the foil.

Stopping power is an atomic property. Bragg and Kleeman [1] found that for a single atom it is proportional to the square root of the mass of the atom, and for a molecule it is proportional to the sum of the square roots of the masses of the atoms composing that molecule. Two films will therefore have the same stopping power if the number of atoms contained per unit area is in the inverse ratio of the square root of their atomic masses.

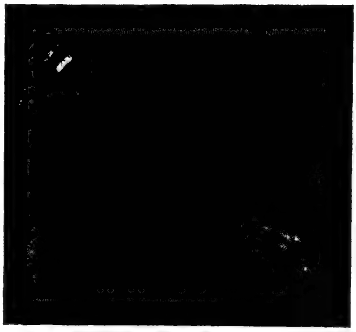


FIG. 6A.

The stopping power of an atom depends upon the speed of the particles. It is, however, almost independent of the speed in the case of substances of about the same atomic weight [2] as air, but it decreases with diminishing speed of the α particle for heavier molecules.

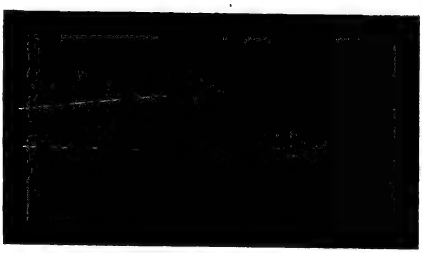


FIG. 6B.

When a stream of particles falls on a sheet of metal foil, some of the particles will come into direct collision with the atoms composing the foil. The particles travel with a velocity of the order of 10^9 cm. per sec., but however fast they travel we should expect that in a direct collision they would be stopped. This, however, was not found to be the case by Bragg and his colleagues. From their researches they concluded that each α particle pursued a rectilinear course, no matter what it encountered; it passed through all the atoms it met, whether they formed part of a solid or a gas, suffering little or no deflection on account of any encounter until very near the end of its course. A thin metal plate placed in a stream of particles robbed every particle of some of its energy, but the number of particles in the stream before and after traversing the thin sheet remained the same. Hence an α particle appeared to pass clean through atoms of matter in its path as if they were not there, in which case two atoms of matter occupy the same space at the same time. In passing through the atom some of the energy of the particle is absorbed and its velocity therefore diminishes as it pursues its course. Further, the slower it moves the more easily is it deviated from its course or scattered.

(iii.) *Visible Tracks of α Particles.*—C. T. R. Wilson [3] succeeded in making visible the tracks of α particles as they pass through a gas. The method consists in suddenly expanding moist air in a closed space, during which operation the moisture condenses on the ions formed by the particle along its path. By suitably illuminating the expansion chamber, the tracks of the particles can be made visible and actually photographed. The tracks of α particles from radium are shown in *Fig.* 6. It will be observed that the tracks are almost all perfectly straight, but there are a few which show abrupt large deflections, especially towards the end of the path (see *Fig.* 6B, which is an enlarged portion of *Fig.* 6A). The chance of a "single" scattering through a large angle increases rapidly with decrease of velocity of the particle, and in consequence such scattering will be most in evidence near the end of the range of the particle.

(iv.) *Counting of α Particles.*—Two methods have been employed to count the number of α particles emitted by a radioactive body.

(*a*) *The Scintillation Method.*—This is based upon the fact that each α particle produces one scintillation when it strikes a screen of a substance such as zinc sulphide (see article on "Luminous Compounds"). The method of procedure is to fix the screen to the microscope at such a distance from the objective that it is perfectly in focus. The microscope

[1] Bragg and Kleeman, *Phil. Mag.*, 1905, x. 318 and 600.

[2] The atomic weight of the hypothetical atom of air may be taken as 14·4.

[3] Wilson, *Roy. Soc. Proc.*, 1912, A, lxxxvii. 277.

is mounted on a graduated stand along which it can slide and be placed at any desired distance from the α ray source. It is advisable to use a microscope with a low-power eyepiece combined with an objective of high light-collecting power, and the area of the field of view need not be larger than about one square millimetre. Also the counting is facilitated if the screen is very faintly illuminated so as to keep the eye the more easily focussed upon it. Measurements can be made most accurately with about 40 scintillations per minute. With more than about 80 or less than about 10, the counting becomes troublesome and uncertain. It is best to count for one or two minutes, afterwards resting the eye to count again for another one or two minutes, and so on. It is also essential to remain in the dark for about half an hour before starting to count, so as to get the eye quite accustomed to the dark.

(*b*) *Rutherford and Geiger's*[1] *Electrical Method.*—In this method the ionisation produced by each α particle is magnified by using the principle of the production of ions by collision. In this way the small ionisation current produced by a single α particle may be magnified several thousand times and thereby be

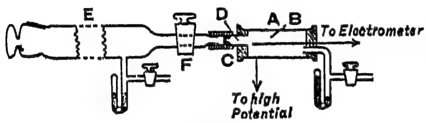


FIG. 7.

sufficiently strong to produce an easily measurable deflection of the needle of an ordinary electrometer. The apparatus employed is shown in *Fig.* 7. The detecting vessel consisted of a brass cylinder A, from 15 to 20 cm. in length and 1·77 cm. in diameter, with an insulated central electrode B connected to one pair of quadrants of an electrometer. The outside of the vessel was connected with the negative pole of a battery giving 1320 volts. The tube D fitting into the ebonite plug C had a circular opening 1·5 mm. in diameter at one end, the opening being covered with a very thin sheet of mica. The thickness of the mica sheet was equivalent in stopping power of the α particle to about 5 mm. of air at ordinary temperature and pressure. The glass vessel E, connected to the detecting vessel as shown, was about 450 cm. in length. The active matter was placed in this vessel in the form of a thin film of small surface area, and its distance from the aperture in D could be varied to any desired value. This vessel was exhausted to a low vacuum whilst the detecting vessel contained carbon dioxide at a pressure of about 4 cm. When the stopcock F was closed no α particles passed into the chamber, and the steadiness of the electrometer needle could thus be tested at intervals during the experiment. On opening the stopcock a small fraction of the total number of α particles emitted by the source passed through the aperture into the detecting vessel. The intensity of the source and its distance from the aperture were adjusted so that three to five α particles entered the chamber per minute. A ballistic throw of the electrometer marked the entrance of an α particle into the chamber.

If Q be the number of α particles expelled per second from the source at a distance r from the aperture of area A, then, since α particles are on the average projected equally in all directions, the number (n) of α particles entering the detecting vessel per second is given by $n = QA/4\pi r^2$. This expression holds so long as each portion of the active source can fire particles through the aperture.

By this method Rutherford and Geiger found that $3{\cdot}57 \times 10^{10}$ α particles were expelled per second from one gramme of radium itself. It is also known that the same number of α particles is emitted per second from one gramme of radium itself and from each of the next three α-ray products in equilibrium with it. So that the number of α particles expelled per second from one gramme of radium in equilibrium with its products is $14{\cdot}3 \times 10^{10}$.

The above method was modified later by the use of a string electrometer, the movements of whose quartz fibre could be photographically recorded on a moving film. In this way it was possible to detect with certainty the effect of each α particle even when 1000 particles

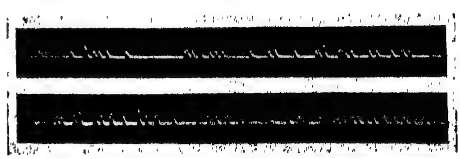


FIG. 8.

entered the detecting chamber per minute. The first method was not accurate when more than ten particles entered the chamber per minute. The records obtained in this way are shown in *Fig.* 8.

Kovarik[2] has devised a method by which the α particles are recorded by a pen on a moving strip of paper. The ionisation current produced by the α particle is magnified by a three-valve audion amplifier. The magnified current is used to operate a very sensitive relay, which in turn operates a local battery circuit accentuating the pen on a chronograph.

[1] Rutherford and Geiger, *Roy. Soc. Proc.*, 1908, A, lxxxi. 141.

[2] Kovarik, *Phys. Rev.*, 1919, xiii. 272.

The sensitive relay operates on a fraction of a milliampere. The method is also applicable to β, γ, and X rays.

(v.) *Ratio of the Charge to the Mass of the a Particle.*—The ratio of the charge to the mass of the α particle can be determined by measuring the deflections produced respectively in a magnetic and in an electric field. It can be shown that the path of a particle of mass m moving with velocity v and carrying a charge e will have a curvature ρ given by the relation

$$H\rho=\frac{mv}{e}, \quad . \quad . \quad . \quad . \quad (7)$$

where H is the strength of the magnetic field. Similarly, when it moves through an electrostatic field produced between two parallel plates whose difference of potential is V,

$$\frac{1}{2V}=\frac{e}{mv^2}. \quad . \quad . \quad . \quad . \quad (8)$$

These two equations enable the values of e/m and v to be determined. Rutherford [1] found by this method that the value of e/m for the α particle was 5070. From theoretical considerations it can be deduced that the value of the ratio of the charge to the mass of the hydrogen atom should be 9647, which is approximately twice that for the α particle. These results may be explained if we assume either that the α particle is a hydrogen molecule or an atom of weight 2 carrying a unit positive charge, or that it is a helium atom (atomic weight 4) carrying two positive charges. To decide between these two alternatives Rutherford and Geiger undertook an accurate measurement of the charge carried by the α particle.

(vi.) *Charge carried by a Particle.*—From the knowledge of the number of α particles emitted by a given product and the total charge carried by these particles, it is possible to deduce the charge carried by a single α particle. The measurement of the charge is, however, made more difficult on account of the fact discovered by J. J. Thomson, that emission of α particles is accompanied by the emission of negatively charged particles, which he termed δ rays. These particles are always emitted when α particles fall on any object. To determine the charge of the α particle the delta rays had to be removed, and this was done by Rutherford and Geiger [2] by placing the radioactive material in a strong magnetic field. The delta particles, which move at low velocities, are thus bent round and return back to the surface which emits them. In this way it was found that the charge carried by the α particle was $9{\cdot}3\times10^{-10}$ e.s.u. The value of the ionic charge had previously been found to be about half this value (the value of the ionic charge at present accepted is $4{\cdot}77\times10^{-10}$ e.s.u.), so that the α particle carried two unit positive charges.

[1] Rutherford, *Phil. Mag.*, 1906, xii. 348.
[2] Rutherford and Geiger, *Roy. Soc. Proc.*, 1908, A, lxxxi. 162.

(vii.) *Nature of the a Particle.*—Evidence so far accumulated pointed to the conclusion that the α particle was a helium atom carrying two unit positive charges. The following experiment carried out by Rutherford and Royds [3] confirmed this.

A quantity of emanation was compressed into a thin-walled glass tube A which was surrounded by a vacuum jacket T (*Fig.* 9). The thickness of the wall was less than 0·01 mm. This was strong enough to withstand atmospheric pressure, and the α particles readily passed through it, as was shown by the scintillations on a zinc sulphide screen held near the tube. The gases in the tube T could be compressed into the spectrum tube S by means of the mercury column H; the colour of the discharge through this tube was then examined spectroscopically. Two days after the emanation was introduced into the tube, the spectrum showed the yellow line of helium, and in six days' time the whole helium spectrum was observed. To show that the helium was not derived by diffusion from the inner tube, the emanation was removed and helium substituted. No trace of helium could be detected in the outer tube after standing for several days, so that the helium in the first instance must have originated from the α particles which had passed through the glass wall of the emanation tube.

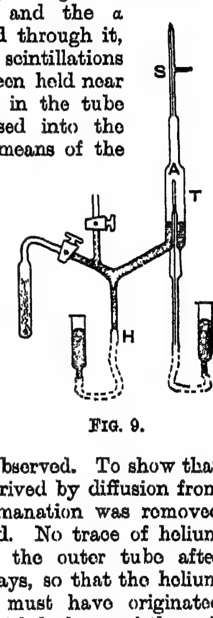


FIG. 9.

There is therefore no doubt that the α particle is a helium atom carrying two positive charges or, in other words, it is a helium atom which has lost two electrons. When its charge is neutralised the α particle becomes a normal helium atom.

§ (12) HYDROGEN PARTICLES.—Before leaving the α particle a brief reference will be made to the recent work of Rutherford, who has made a close quantitative study of the effect on an α particle of its passage through atoms of matter. Bragg had shown that the α particles pass straight through the atoms of matter in their path; this is undoubtedly the case with the majority of the α particles in the stream; but there are exceptions, in which cases the α particle suffers large deflections. This "occasional large-angle

[3] Rutherford and Royds, *Phil. Mag.*, 1909, xvii. 281.

scattering" is to be distinguished from the very slight deviations in different directions according to the laws of probability which later more detailed examination has shown to happen to the α particle as it traverses matter. These facts point to the conclusion that the central nucleus of the atom is very small, but however small it is an occasional α particle is certain to strike it absolutely "head on." It is these occasional close encounters that have been examined by Rutherford.[1] When such a collision takes place between the α particle and the nucleus of a heavy atom, the former will either be violently swung out of its path if the collision is not absolutely "head on," or repelled the way it came almost at its original velocity if it is "head on."

When, however, a "head-on" collision takes place between the α particle and the nucleus of a lighter atom—for example, atoms of hydrogen—then, in this case, the hydrogen atom would be propelled in the same direction as that of the α particle with a velocity far greater than the velocity of the original particle, and it would consequently travel a greater distance in hydrogen gas than the α particle before it is stopped. This was actually found to be the case in hydrogen. These high-velocity atoms were, for the sake of clearness, termed by Rutherford H-particles. The same phenomenon was observed also in the case of oxygen and nitrogen. In the case of nitrogen, however, there were observed in addition to the N-particle, whose range was only slightly longer than that of the α particle itself, particles of long range and other characteristics exactly similar to the H-particles produced in hydrogen gas, one H-particle being observed for every twelve N-particles produced, which suggests that the nucleus of the nitrogen atom struck by an α particle is occasionally shattered by the collision and that hydrogen atoms are produced from it. Later experiments carried out on these particles by subjecting them to the action of electric and magnetic fields proved that the H-particles generated in hydrogen and also in nitrogen consisted of hydrogen atoms, each carrying a single positive charge. The "N-particles" and "O-particles" were, however, not singly charged atoms of nitrogen and oxygen respectively, but were found to be identically the same and consisted of an entirely new particle of mass 3 carrying two positive charges.

Hence a nitrogen atom coming into close nuclear collision with an α particle is shattered, yielding in some cases atoms of hydrogen of mass 1 carrying a single positive charge, and in others atoms of a new kind of mass 3 carrying a double positive charge. In the case of oxygen only atoms of the new kind of mass 3 are produced.

§ (13) β Rays. (i.) *General Properties.*—The general properties of the β rays are identical with those of the cathode rays in a discharge tube. They are negatively charged particles or "electrons" projected at high velocities; the velocity of a β particle is far higher than that of the fastest known cathode ray, in some cases the velocity is almost indistinguishable from that of light. It is probable that the β ray has its origin in the central positively charged nucleus of the atom and, as in the case of the α particle, each disintegrating atom emits one β particle only.[2]

The character of the β rays is brought out very clearly in Strutt's[3] radium clock. A small quantity of radium salt is enclosed in the tube A, whose walls are thick enough to absorb the α rays (*Fig.* 10). This is suspended in an outer tube and insulated from it. A pair of gold leaves are attached to the bottom of the tube A and are in metallic connection with the radium. If the outer tube is highly exhausted, the radium becomes positively charged owing to the loss of β particles and the leaves diverge. If the vacuum is high several hundred volts may be reached before the loss of charge through the insulator and gas balances the rate of supply. In one form of the instrument, two plate electrodes B and C are sealed into the glass bulb, one on either side of the leaves and both connected to earth. These serve to discharge the leaves after a certain divergence has been reached. The charging and discharging of the leaves will in this way go on indefinitely.

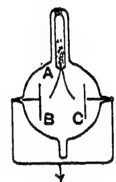


Fig. 10.

(ii.) *Magnetic Spectrum.*—β rays are readily bent by comparatively weak magnetic fields, and, since they carry a negative charge, the deflection is in the opposite direction to that produced in the case of the α rays. If all the β rays in a beam emitted by a radioactive body were travelling at the same velocity, the deflection produced by the magnetic field would be the same for each ray; this, however, is never found to be the case, so that the β particles are not expelled from the radioactive body all with the same velocity. In some cases the magnetic spectrum is continuous, in other cases there are distinct lines in the spectrum showing that there are present in the beam groups of rays of definite velocity.

An example of the spectrum of the β particles emitted by the active deposit of

[1] Rutherford, *Phil. Mag.*, 1919, xxxvii. 537; *Roy. Soc. Proc.*, 1920, A, xcvii. 374.

[2] Moseley, *Roy. Soc. Proc.*, 1912, A, lxxxvii. 230.
[3] Strutt, *Phil. Mag.*, 1903, vi. 588.

radium, photographed by Rutherford, is shown in *Fig.* 11. The rays, after passing through a narrow slit, fell on a photographic plate placed at an angle of about 45° with the horizontal. The central line is due mainly to the α rays; the β rays show on both sides of this line because the magnetic field was reversed during the experiment. The lines in some cases are very numerous; for instance, 48 lines for radium C have been observed with velocities ranging up to 0·986 of the velocity of light. It is possible, however, that the initial energy of each α particle from a single radio-element is the same. The particle has its origin in the nucleus, and it has to pass through successive rings of electrons surrounding the nucleus before it leaves the atom. In this process energy may be lost in quanta, and there is evidence to show that this is actually the case. The energy lost in this way reappears again as γ ray energy.

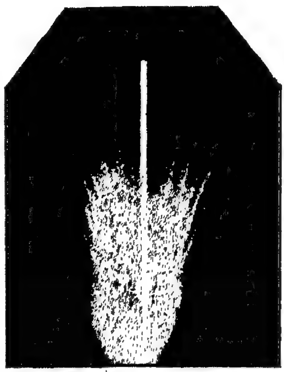


FIG. 11.

(iii.) *Absorption of β Rays by Matter.*—The absorption of β rays is investigated by placing thin sheets of the absorbing material in the path of the rays and measuring the activity through different thicknesses. It is found that the absorption of the rays follows an exponential law, that is, if I_0 is the activity when no absorbing material is interposed and I the activity when the rays pass through a thickness d, then

$$I = I_0 e^{-\mu d}, \quad . \quad . \quad (9)$$

where μ is known as the coefficient of absorption of the β rays in the absorbing material. The absorption coefficients of the β rays from uranium X in aluminium, copper, and lead are 14 cm.$^{-1}$, 66 cm.$^{-1}$, and 103 cm.$^{-1}$ respectively.

Since the rays are absorbed according to an exponential law, then each equal successive thickness of material interposed produces the same percentage reduction in the radiation remaining after traversing the previous layers. It was initially thought that this exponential law of absorption was a proof that all the rays travelled with the same velocity—that is, that the beam was homogeneous. Wilson [1] showed, however, that this was no test for homogeneity of the rays. The experiments of Von Baeyer, Hahn, and Meitner [2] also showed that β rays which obey the exponential law of absorption were divided into groups of different velocities when subjected to a magnetic field. When the differences of velocities of the groups are large, then the law of absorption is no longer exponential.

To obtain the true value of the absorption coefficient of β rays it is necessary to work with preparations, such as active deposits, that can be obtained free from other matter, in the form of very thin films in which the absorption of the β rays is negligible. The β rays starting from the surface of a layer of uranium oxide, for instance, will have a higher velocity than those that come from beneath the surface. The effect on the emission of β particles from different thicknesses of uranium oxide is shown in *Fig.* 12. The ionisation is proportional to the thickness for thin layers, but as the thickness becomes greater it increases less rapidly, ultimately reaching a steady value, in which case the radiation from the lower layers is completely absorbed by the active material in the upper layers. The curve gives sufficient data to calculate the absorption coefficient of the β rays in the uranium oxide itself. Considering a thin layer of thickness dx at a distance x below the surface of the active material, we have

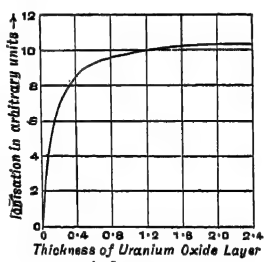


FIG. 12.

$$dI = I_0 e^{-\mu x} dx, \quad . \quad (10)$$

where dI is the amount of radiation entering the measuring vessel from the layer dx of active material. Hence the total amount of radiation emitted by a layer of finite thickness d is given by the expression

$$I_d = I_0 \int_0^d e^{-\mu x} dx = \frac{I_0}{\mu}(1 - e^{-\mu d}). \quad . \quad (11)$$

As the thickness of the layer increases, this tends to the value I_0/μ, which is a constant, I_0 being the intensity of the radiation emitted by a very thin layer of the active material, in which case there would be no absorption, and μ is the absorption coefficient of the radiation

[1] W. Wilson, *Roy. Soc. Proc.*, 1909, A, lxxxii. 612.

[2] Von Baeyer, Hahn, and Meitner, *Phys. Zeits.*, 1911, xii. 273 and 378; 1912, iii. 264.

in the material. Indicating the steady value of the intensity by I_∞, then

$$I_d = I_\infty(1 - e^{-\mu d}). \quad . \quad . \quad (12)$$

The value of the coefficient of absorption of the uranium oxide for the β rays can be calculated from this expression by substituting the values of I_d and I_∞ obtained from the curve.

(iv.) *Scattered β Rays.*—When β rays fall on a layer of matter, part of the radiation is absorbed and part is scattered. This scattered radiation is identical in type with the primary β rays, but is always of less average penetrating power than the latter. Its amount increases as the atomic weight of the radiator increases. Kovarik and Wilson [1] have also shown that the amount of scattered radiation from different materials increases with the velocity of the β rays up to a certain point and afterwards decreases. The above results apply to thick layers of material in which a considerable amount of absorption takes place. In order to be able to compare experimental results with any theory of scattering, it is desirable to do away with appreciable absorption and consequently to use very thin layers of scattering material. Two theories of scattering have been put forward, the one by J. J. Thomson [2] and the other by Rutherford.[3] On the former theory the scattering is supposed to be due to the chance combination of a multitude of successive minute deflections. The β particle as it encounters one atom after another in its passage through matter is deflected in different directions, but the average value of this deflection over a very large number of encounters will be a finite quantity. If θ is the average deflection due to each encounter with an atom, it can be shown that the mean deflection after n encounters is $\theta\sqrt{n}$. Hence, if the rays pass through a plate of thickness t containing N atoms of radius b per unit volume, the average deflection which they experience is $\theta\sqrt{N\pi b^2 t}$.

Rutherford, on the other hand, assumes that a moving electrified particle can be deflected through a large angle by a single atomic encounter, and he attributes the main features of the scattering of particles when they pass through thin layers of matter to this single scattering. The two theories differ also as to the constitution of the atom—in the former the atom is supposed to be constituted of N_o electrons accompanied by an equal quantity of positive electricity, the latter being distributed with uniform volume density throughout a sphere whose volume is equal to that of the atom; whilst in the latter the atom is imagined to consist of a central positive nucleus carrying a charge $+N_o e$ surrounded by N_o electrons carrying altogether a negative charge equal to the positive charge on the nucleus.

When the layer of scattering material is so thin that there is no absorption of the radiation, both theories lead to constant values for the fractions $\phi/\sqrt{t_o}$ and $mu^2/\sqrt{t_o}$, where ϕ is any particular mean deflection, t_o the thickness of material required to cut the radiation down to half value, and $\frac{1}{2}mu^2$ the energy of the β particle. The actual values of the constants are different, however, on the two theories, and consequently the values of the number of electrons in the atom which can be deduced from the value of $\phi/\sqrt{t_o}$ will differ in the two cases.

The experiments of Crowther [4] on the scattering of β rays by very thin sheets of different materials showed that the fractions $\phi/\sqrt{t_o}$ and $mu^2/\sqrt{t_o}$ were both constant in accordance with theory. Calculating the number of electrons in atoms of the metals Al, Cu, Ag, and Pt, which were employed in the investigation, the results based on Rutherford's theory agree more closely with the number of electrons per atom as determined by X rays than do those based on Thomson's theory.

§ (14) γ Rays. (i.) *General Properties.*—One of the chief characteristics of these rays is their great penetrative power. It is possible by means of the electroscope to detect the γ radiation emitted by 30 milligrammes of radium after it has passed through 30 cm. of iron. The γ rays are always found associated with β rays, and they are emitted in large amounts only from those radioactive bodies that emit penetrating β rays. A beam of β radiation falling on any substance gives rise to γ radiation in the same way that a beam of cathode rays produces X rays. Just as X rays liberate corpuscular radiation when they fall on a substance, so γ radiation liberates β particles, the only difference being in the velocity of the particle liberated.

(ii.) *Absorption of γ Rays from Radium.*—To measure the absorption of γ rays in any substance it is necessary to eliminate the effect due to the β rays. The latter may be removed either by deflecting them from the measuring apparatus by means of a magnetic field or by absorbing them in a layer of matter of sufficient thickness. The latter method is the one usually adopted, a screen of lead 3 mm. thick being sufficient to absorb almost completely the β rays emitted by radium and its products.

[1] Kovarik and Wilson, *Phil. Mag.*, 1910, xx. 866.
[2] J. J. Thomson, *Camb. Phil. Soc. Proc.*, 1910, xv. 465.
[3] Rutherford, *Phil. Mag.*, 1911, xxi. 669.
[4] Crowther, *Roy. Soc. Proc.*, 1910, A, lxxxiv. 226.

The absorption of the rays can be measured by means of an electroscope. Sheets of known thickness of the material investigated are placed in the beam, and from the knowledge of the deflection of the leaf with and without the absorbing material respectively in front of the electroscope the absorption coefficient of the material for the radiation can be calculated by means of the formula

$$I = I_0 e^{-\mu t}, \quad . \quad . \quad (13)$$

where I and I_0 are the intensities of the radiation entering the electroscope with and without the absorbing screen, t the thickness of the screen, and μ the coefficient of absorption.

It is important that the walls of the electroscope are sufficiently thick to absorb β radiation that may be excited by the radiation which falls on objects in the vicinity; also the windows through which the gold leaf is viewed should be well shielded. The values of the absorption coefficient depend somewhat on the arrangement of the apparatus. The following figures for the absorption of γ rays from radium in lead are due to Tuomikoski: [1]

Table II

Thickness of Lead in cm.	Ionisation Current.	Thickness of Lead in cm.	Ionisation Current.
0·3	100	7·0	2·57
1·0	61·6	8·0	1·62
2·0	33·1	9·0	1·00
3·0	19·9	10·0	0·63
4·0	11·7	11·0	0·39
5·0	7·07	12·0	0·30
6·0	4·26	..	..

The values of the absorption coefficients over the different thicknesses are given in the following table:

Table III

Thickness of Lead in cm.	μ cm.$^{-1}$.
0·4– 1·0	0·70
1·0– 2·2	0·58
2·2– 5·4	0·52
5·4–12·0	0·50

The absorption of γ rays in various substances shows that for substances whose densities lie between 2·6 and 8·8 the absorption is very approximately proportional to the density. The value of the mass absorption coefficient (μ/ρ) is therefore constant over this range of densities; for other substances the densities of which do not lie within the above limits the value of μ/ρ is greater.

γ rays from different radioactive products have different penetrating power. The relative values of the absorption coefficients are given in the following table: [2]

Table IV

Absorbing Screen.	γ Rays from			
	Radium C.	Thorium D.	Mesothorium 2.	Uranium X.
Lead . . .	1·0	0·924	1·24	1·45
Zinc	1·0	0·82	1·06	1·18
Paraffin wax .	1·0	0·78	1·26	1·08

The rays from thorium D are more penetrating, while the rays from mesothorium 2 and uranium X are slightly less penetrating than the rays from radium C.

(iii.) *Scattered γ Radiation.*—In addition to the production of β rays, γ rays in traversing matter give rise to a scattered radiation of the γ type. The amount of this scattered radiation appearing on the emergent side of a radiator is always much greater than that on the incident side. This asymmetry in the distribution of the radiation is also observed with X rays, but the effect is not so marked as with γ rays. Florance [3] found that the quality of the scattered radiation varied at different angles with the direction of the primary beam, the radiation scattered through a large angle being much more easily absorbed than the primary γ rays. The scattered radiation appearing on the side of incidence is also softer than that on the emergent side of the radiator.

(iv.) *Nature of γ Rays.*—There has been a good deal of controversy as to the nature of γ rays, but recent experiments on the diffraction of the rays have definitely proved that the rays are of the same nature as X rays, *i.e.* aetherial pulses, the only difference being that the wavelength of γ radiation is much shorter than that of X radiation.

Rutherford and Richardson [4] showed that radium C emits one and radium B two types of γ radiation, each of which is exponentially absorbed in aluminium. These radiations were carefully examined by Rutherford and Andrade [5] by reflecting them at a face of rock salt crystal. The source of radiation was a thin-walled α ray tube containing about 100 millicuries of emanation in equilibrium with its products A, B, and C. A diverging cone

[1] Tuomikoski, *Phys. Zeits.*, 1909, x. 372.

[2] Russell and Soddy, *Phil. Mag.*, 1911, xxi. 130.
[3] Florance, *ibid.*, 1910, xx. 921.
[4] Rutherford and Richardson, *ibid.*, 1913, xxv. 722.
[5] Rutherford and Andrade, *ibid.*, 1914, xxvii. 854; xxviii. 263.

of rays fell on the crystal face, and the distribution of the reflected radiation was examined by the impression produced on a photographic plate placed 10 cm. from the centre of the crystal.

The spectrum of the γ radiation from Ra B was found to consist of 21 lines having wave-lengths ranging from $0{\cdot}793 \times 10^{-9}$ cm. to $1{\cdot}365 \times 10^{-8}$ cm., the two strongest lines being reflected from the (100) face of rock salt at the angles 12°·05 and 10°·05, and therefore having wave-lengths $1{\cdot}176 \times 10^{-8}$ cm. and $0{\cdot}983 \times 10^{-8}$ cm.

The wave-lengths of the penetrating γ radiation from Radium C were found to range from $0{\cdot}71 \times 10^{-9}$ cm. to $1{\cdot}96 \times 10^{-9}$ cm.

§ (15) Various Effects produced by the Radiations.—It has already been pointed out that the radiations ionise gases through which they pass and, in consequence, make them temporarily conductors. This is effectively demonstrated by bringing a little radium near the secondary terminals of an induction coil. If the two terminals are so separated that a spark just does not pass between them when the coil is running, by bringing some radium near the gap the spark will readily pass. This property of the radiation from radium is made use of when it is desired, during the course of an investigation, to prevent electric charges accumulating on the surfaces of different parts of the apparatus.

The salts of radium are all luminous in the dark, and the radiations produce marked luminescence in certain salts (see article on "Radium"). Certain bodies after exposure to the radiations become luminous when they are heated to a temperature much below that required to produce incandescence. Fluor-spar and kunzite possess this property of thermo-luminescence to a marked degree. These substances are able to store the energy which they take up for a long period of time. The explanation suggested for this effect is that the rays cause chemical changes which are permanent until heat is applied which releases in the form of visible light the energy absorbed.

Glass exposed to the radiations for some time becomes coloured, soda glass is coloured a deep violet, and after a long period under the influence of the rays it becomes almost black. Other glasses are coloured yellow and brown under the action of the rays. This coloration is produced by all three types of radiation; that due to the α rays extends only a short distance corresponding to the range of the α particles in the glass, whilst that due to the β and γ rays extends throughout the whole of the glass. Mica plates acquire a brown or a black colour under the action of α rays. It had long been observed that certain kinds of mica contained small coloured areas whose sections were always circular and in whose centre there was usually a minute crystal of foreign matter. Under the action of polarised light these exhibited the property of pleochroism, and for this reason were called "pleochroic halos." Joly [1] found that the nucleus of these areas was radioactive and that the coloration was due to the α rays expelled from this nucleus. The halos exhibit a well-marked structure clearly shown in *Fig.* 13, which is a micro-photograph of a halo whose nucleus is a uranium mineral.[2] The central dark area is produced by the collective action of all the α particles emitted by the nucleus, and the boundary of it defines the range of the α particles from radium itself. The next dark edge corresponds to the range of the swifter α particles from radium A, and the edge of the outer ring to the range of the α particles from radium C. It is possible to find in the mica halos in various stages of development. In some may be found only the central "pupil," 0·013 mm. in radius corresponding to the range of the slower α particles from radium, but in more developed cases an outer ring always appears whose radius is 0·03 mm. corresponding to the range in mica of the α particles from radium C. Other halos have different diameters, and it is possible to decide from the diameter of the halo whether the nucleus contains uranium or thorium mineral. The fastest α particle emitted in the thorium series would travel a distance of 0·038 mm. in mica, and it was found on examination that halos of this dimension were obtained when the nucleus consisted of thorium mineral.

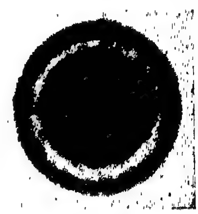


Fig. 13.

It has been estimated that the number of α particles emitted by the nuclei of these halos is of the order of 100 per year. The effect of this expulsion on the mica would be infinitesimal over a short period, so that it has probably taken several hundred million years to produce the actual halos that have been observed.

The action of α rays on photographic films examined by Kinoshita and Ikeuti [3] are of interest in this connection. A sewing needle carrying at its point a minute trace of radium active deposit was employed as a source of α rays. After the point of the needle had been in contact with the photographic plate for a short time, the plate was developed and a fine spot became visible to the naked eye. Under the microscope this spot was found to

[1] Joly, *Phil. Mag.*, 1907, xiii. 381.
[2] Joly and Fletcher, *ibid.*, 1910, xix. 630.
[3] Kinoshita and Ikeuti, *ibid.*, 1915, xxix. 420.

consist of a multitude of radial trails of silver grains around a circular dark nucleus. Halos were also observed and were shown to be produced by a process similar to that of the pleochroic halos.

Another important effect of the radiations is the decomposition of water. This is brought about by all three types, but the effect due to the α rays is much greater than that due to the β and γ rays. The effect is probably due to the ionisation of the water molecule, which is accompanied by the chemical dissociation of the molecule with the liberation of hydrogen and oxygen. Hence it is important, when sealing radium in tubes, to see that the salt is dry, otherwise in the course of time the pressure of the evolved gases inside the tube may become so great as to burst it.

§ (16) The Emanations. — Rutherford[1] found that thorium emitted, in addition to the radiations, a radioactive gas which he called the emanation. The amount given off is exceedingly small, but the ionisation produced by it is sufficiently large to be readily measured by means of an electroscope. The emanation can be carried away from the neighbourhood of the substance emitting it by a current of air, and its activity will not suffer appreciable loss when it is passed through a plug of cotton-wool or bubbled through solutions on its way to the measuring apparatus.

Shortly after the discovery of thorium emanation, both radium and actinium were found to give off emanations. The activity of each of the emanations decays according to an exponential law with time. Actinium and thorium emanations are short-lived, their half-value periods being 3·9 seconds and 54 seconds respectively, and on this account it is difficult to remove them quickly enough from solution before they disintegrate. Radium emanation, on the other hand, has a half-value period of 3·85 days, and practically the whole of it can be removed by passing a stream of gas through, by exhausting, or by boiling the solution.

Solid preparations differ enormously in the readiness with which they part with their emanation. In the case of thorium the best emanators are the hydroxide and the carbonate. After ignition to oxide the emanating power is greatly reduced. The higher the temperature of ignition the greater is this reduction, until after ignition at a white heat the emanating power may be reduced to a small fraction (a few per cent) of that of the unignited compound. The nitrate, oxalate and sulphate are all poor emanators. In the case of radium, the bromide is a better emanator than the chloride. From the sulphate and, to a less extent, the carbonate, very little emanation escapes in the solid. It has been found, however, that the same compounds may exhibit in this respect a marked variation in behaviour.

[1] Rutherford, *Phil. Mag.*, 1900, xlix. i.

Hence if the radioactive substance is required to retain its emanation in the open it is best in the case of radium to keep it as sulphate, and in the case of the other two radio-elements in the form of highly ignited oxides. If the emanation is required, the radioactive material is best kept in solution.

All the three emanations have distinct chemical and physical properties; they have characteristic bright line spectra, they are also chemically inert, and in consequence they are classed with the argon-helium group of the periodic table. They also all emit α particles and can be condensed at liquid air temperature.

§ (17) Radium Emanation.—If a solution of radium is kept in a sealed flask, the emanation steadily accumulates with time, reaching the equilibrium value in about a month's time. Oxygen and hydrogen are at the same time evolved owing to the decomposition of the water by the radium. On exploding the gases, a slight excess of hydrogen remains with the emanation. The hydrogen can be removed by passing the mixture into a bulb immersed in liquid air, where the emanation will be condensed and the hydrogen together with any helium produced by the emanation can then be pumped off. The liquid boils at about $-62°$ C. at ordinary atmospheric pressure and the solid melts at about $-71°$ C. The liquid emanation is colourless and transparent, whereas the solid is opaque and glows with great brilliancy and is of a steel-blue colour. As the temperature is still further lowered, the colour changes to yellow and becomes a brilliant orange red at the temperature of liquid air.

The volume of emanation at normal temperature and pressure, in equilibrium with 1 gram of radium, is found by experiment to be 0·62 cubic mm. This is also the value calculated on the assumption that the emanation molecule consists of single atoms, so that the agreement between theory and experiment directly proves the monatomic nature of the molecule. The atomic weight of radium emanation is 222·0, four units less than that of its parent, radium, which has lost an α particle in the process of disintegration.

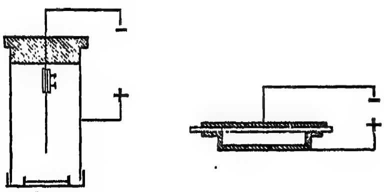
Radium emanation is soluble in water, the coefficient of solubility being about 0·3 at ordinary temperature and 0·12 at 80° C. It is more soluble in petroleum and toluene, the solubility coefficients being 9·5 and 11·7 respectively at ordinary temperature.

§ (18) Active Deposits.—Bodies which have been exposed for some time to the emanations from radium, thorium, and actinium acquire a temporary activity of their own, owing to the deposition of active matter on their surfaces. This "active deposit" consists of the dis-

integration products of the emanation which, in each case, are non-volatile at ordinary temperatures and have a relatively short period of life. When the emanations are allowed to decay in a strong electric field Rutherford found that the activity was confined entirely to the negative electrode, which indicates that the carriers of the active material must be positively charged. Owing to the fact that these products are produced from the emanation by the expulsion of α particles, it would be expected that the active deposit would be negatively charged. This apparent anomaly may be explained as follows. The atom emitting the α particle recoils and collides with the gas molecules in its path. During these collisions negative electrons are shaken out of the recoiling atom, which in consequence acquires a positive charge. Evidence in support of this view is the fact that α particles emitted by an active material are accompanied by a large number of δ particles which, as we have already seen, are slow speed negative electrons.

The active deposit may be concentrated and collected on a metal surface or a wire by making the latter the negative electrode inside a metal vessel containing the emanation, the vessel itself being connected to the positive pole of the source of potential. The potential gradient necessary to concentrate the whole of the active deposit on the negative electrode varies according to circumstances. It will, in general, be the same as that necessary to produce "saturation" of the ionisation current through the gas under the conditions of experiment. Higher potential gradients than 50 volts per centimetre are seldom necessary except in the case of intensely radioactive preparations or when the area of the surface to be covered with the deposit is very small.

To collect thorium and actinium active deposits the arrangements shown in *Fig.* 14 are



(a)

FIG. 14.

convenient for concentrating the deposits on a wire and on a plate respectively. The active material is placed in a shallow dish in each case, and the emanation escaping from it allowed to diffuse into the vessel. As it decays it produces active deposit which is driven by the electric field to the cathode, a wire in the one case and a plate in the other. To obtain the maximum activity the surfaces should be exposed for three hours in the case of actinium and three days in the case of thorium. It is, of course, advisable to use the salts that are the best emanators in each case. It is also an advantage to keep the salt slightly damp.

The above arrangements are not suitable in the case of radium on account of the long life of the emanation and its tendency to remain occluded in the radium salt. In this case the radium salt is dissolved in hydrochloric acid. The solution is placed in a stoppered bottle and kept positively charged by means of a wire which dips into it. The surface to be coated with active deposit is suspended above the solution in the bottle and is charged negatively. An exposure of three hours suffices to collect the maximum amount of active deposit.

In dealing with large quantities of radium the above method is unsuitable, because the stopper has to be removed when the surface to be coated is inserted into the bottle, with the result that an appreciable amount of emanation is lost. Also part of the emanation remains in solution and does not contribute to the activation of the surface. The following method, which diminishes the risk of any loss and makes the best use of the radium, is always employed when large quantities of radium are dealt with (*Fig.* 15).

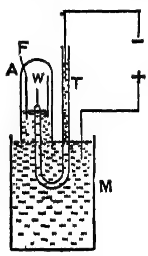


FIG. 15.

The emanation is first of all collected over mercury in the glass tube A. This is then introduced into the mercury reservoir M. The wire w which is to be coated with active deposit is sealed into the bent glass tube T and is connected to the negative pole of the battery. It is surrounded by an iron sheath F which fits the tube A and is in metallic contact with the mercury which is connected to the positive pole of the battery. It is advisable to use a high potential gradient in order to collect as much of the active material as possible—even 1000 volts per centimetre may be necessary in some cases. It is also preferable to use a platinum wire on which to collect the active deposit, as it is often necessary to dissolve the latter in strong acid or to volatilise it by exposure to a high temperature.

§ (19) RADIOACTIVE RECOIL.—Since the mass of the α particle is equal to that of a helium atom ($6{\cdot}56 \times 10^{-22}$ gm.), and it is expelled at a high velocity from the atom, it possesses considerable momentum. At the moment the α particle is expelled the residue

of the atom acquires an equal and opposite momentum and recoils at a considerable velocity. The recoiling atom is, therefore, able to penetrate a certain thickness of matter before it is brought to rest, and may even leave the surface upon which the radioactive product is deposited and be concentrated on another surface by the application of an electric field. It is possible in this way to separate certain disintegration products from the parent atoms.

Consider, for example, a surface made active by a short exposure to radium emanation. Immediately after removal from the emanation the activity is almost entirely due to radium A. An atom of radium A disintegrates with the expulsion of an α particle whose velocity is $1{\cdot}77 \times 10^9$ cm. per second. The atomic weight of radium B—the residual part of an atom of radium A after the expulsion of an α particle—is 214, so that by the principle of momentum its velocity of recoil will be $3{\cdot}27 \times 10^7$ cm. per second. An atom of radium B therefore leaves the active material with this velocity, but since its kinetic energy is small (compared with that of the α particle) it will be able to penetrate only a very small thickness of matter—it will be stopped by one-tenth of a millimetre of air at atmospheric pressure. A recoiling radium D atom from radium C has a "range" of 0·14 mm. in air at atmospheric pressure. Since the range of the recoil atoms is so short, it is necessary for successful separation by recoil to obtain the active material in the form of a thin uniform deposit on a carefully polished surface. This is supported in a good vacuum opposite to the surface on which it is desired to collect the recoil product, and connected to the positive pole of a battery. Almost all the recoil atoms moving away from the active plate, which is half the total number, may be separated in this way.

§ (20) Theory of Successive Transformations.—A full description of the theory of successive transformations of radioactive matter will not be attempted here. A few simple cases only will be considered which will illustrate the application of the theory to practice.[1]

(i.) *Decay of Thorium Active Deposit.*—The active deposit of thorium consists of the members of the thorium series included below thorium emanation in Table I., and of these only the first three need be taken into account when making α ray measurements. The products thorium A and thorium C emit α rays whilst thorium B emits soft β rays. Further the period of thorium A is only 0·2 second, so that the active deposit after removal from the emanation will behave as if it consisted only of two substances, namely thorium B and thorium C.

(a) *Short Exposure to Emanation.*—In this case the plate to be made active is exposed to the emanation for a few minutes and then removed. The plate is then transferred into an α-ray electroscope; in the course of this transference, unless it is carried out exceedingly rapidly, the thorium A is completely converted to thorium B. Assuming that the thorium A has all changed into thorium B by the time observations are commenced, the activity of the deposit at the beginning is due entirely to β rays from thorium B, and consequently is very feeble. As the thorium B disintegrates with the production of thorium C, an α-ray product, the activity rapidly rises and reaches a maximum value in about four hours (see curve ABC, *Fig.* 16). It then

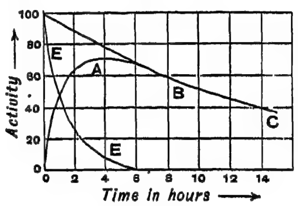


Fig. 16.

decreases and finally decays exponentially with the period of thorium B.

The rate at which the activity varies with time can be calculated if it be assumed that the influence of thorium A is negligible and that initially there is only thorium B present. This product disintegrates in accordance with Table I., thorium B into thorium C, and thorium C into thorium D.

Let P and Q be the number of atoms of the products B and C respectively present at a certain instant, then the rate of increase of thorium C will be given by the difference between the amount of thorium C produced by thorium B and the amount of thorium C disintegrating into thorium D, so that

$$\frac{dQ}{dt} = \lambda_1 P - \lambda_2 Q, \quad . \quad . \quad . \quad (14)$$

where λ_1 and λ_2 are the radioactive constants of B and C respectively, and P is given by the equation

$$P = P_0 e^{-\lambda_1 t}, \quad . \quad . \quad . \quad (15)$$

where P_0 is the initial number of atoms of thorium B present. The amount of thorium C present initially is zero, so that when $t=0$, $Q=0$, and the solution of equation (14) may be shown to be

$$Q = \frac{P_0 \lambda_1}{\lambda_2 - \lambda_1} \cdot [e^{-\lambda_1 t} - e^{-\lambda_2 t}] \quad . \quad (16)$$

[1] See Makower and Geiger, *Practical Measurements in Radioactivity*, p. 80.

This equation should give the same curve as ABC in *Fig.* 16, when the correct values of λ_1 and λ_2 are substituted in it.

The experimental curve shows a maximum, and the time at which this maximum occurs, obtained from (16), is given by

$$t_m=\frac{1}{\lambda_1-\lambda_2}\cdot\log\frac{\lambda_1}{\lambda_2}. \quad . \quad . \quad (17)$$

The values of the radioactive constants λ_1 and λ_2 may be deduced from the experimental curve ABC. The end portion BC of this curve follows a simple exponential law, the decay being governed by the period of the product of longer life. Hence the value of one of the transformation constants is immediately obtained. From equation (16) the curve is made up of two exponentials. Extrapolating the curve BC back to zero and subtracting the experimental curve from it, another exponential curve is obtained, represented in the figure by curve EE. This gives the value of the second transformation constant. To decide to which product each of the constants is to be assigned, it would be necessary to separate the two products and examine the rate of decay of activity for each separately.

The theory of a short exposure to actinium emanation is the same as that for thorium, but in this case the maximum occurs after nine minutes, and in consequence exposure to the emanation must not last for more than a fraction of a minute.

(*b*) *Long Exposure to Emanation.*—It will be assumed here that the exposure is so long that the products have had time to reach a state of radioactive equilibrium. If n_0 be the number of atoms of thorium A deposited per second from the source, then

$$n_0=\lambda_1P_0=\lambda_2Q_0=\lambda_3R_0, \quad . \quad . \quad (18)$$

where P_0, Q_0, and R_0 are the quantities of the products B, C, D present when equilibrium has been reached.

Combining equation (18) with equations (14) and (15) we obtain the equation

$$Q=\frac{\lambda_1P_0}{\lambda_2-\lambda_1}\left[e^{-\lambda_1t}-\frac{\lambda_1}{\lambda_2}\cdot e^{-\lambda_2t}\right], \quad . \quad (19)$$

which gives the number of atoms of thorium C present at any time after the termination of the exposure. This curve is again made up of two exponentials, and the values of λ_1 and λ_2 may be deduced in a similar manner to that described above.

It should be mentioned that the same activity curves would be obtained with thorium and actinium if the activity were measured by means of the β rays instead of the α rays as above.

(ii.) *Decay of the Active Deposit of Radium.*—It will be observed from Table I. that the period of radium A is comparable with the periods of the products which follow it, and on this account the case of the decay of the active deposit of radium presents more difficulty than that of thorium or actinium.

Supposing that there are P, Q, and R atoms of radium A, radium B, and radium C present at a certain time t, the amount of each will vary according to the following equations:

$$\left.\begin{aligned}\frac{dP}{dt}&=-\lambda_1P,\\ \frac{dQ}{dt}&=\lambda_1P-\lambda_1Q,\\ \frac{dR}{dt}&=\lambda_2Q-\lambda_3R.\end{aligned}\right\} \quad . \quad . \quad (20)$$

In the case of a *short exposure* to the emanation, when only radium A will have had time to be deposited, $Q_0=R_0=0$, when $t=0$, and the above equations become

$$\left.\begin{aligned}P&=P_0e^{-\lambda_1t},\\ Q&=\frac{\lambda_1P_0}{\lambda_2-\lambda_1}\cdot\left[e^{-\lambda_1t}-e^{-\lambda_2t}\right],\\ R&=\lambda_1\lambda_2P_0\left[\frac{e^{-\lambda_1t}}{(\lambda_2-\lambda_1)(\lambda_3-\lambda_1)}+\begin{matrix}\text{two similar}\\ \text{terms}\end{matrix}\right].\end{aligned}\right\}(21)$$

In the case of a *long exposure*, the initial conditions, when the deposit is removed from the emanation, are

$$\lambda_1P_0=\lambda_2Q_0=\lambda_3R_0, \quad . \quad . \quad (22)$$

whence

$$\left.\begin{aligned}P&=P_0e^{-\lambda_1t},\\ Q&=Q_0\left[\frac{\lambda_2}{\lambda_2-\lambda_1}\cdot e^{-\lambda_1t}+\frac{\lambda_1}{\lambda_1-\lambda_2}\cdot e^{-\lambda_2t}\right],\\ R&=R_0\left[\frac{\lambda_2\lambda_3e^{-\lambda_1t}}{(\lambda_2-\lambda_1)(\lambda_3-\lambda_1)}+\begin{matrix}\text{two similar}\\ \text{terms}\end{matrix}\right].\end{aligned}\right\}(23)$$

These equations enable the numbers of atoms of radium A, radium B, and radium C present at any time after a short or a long exposure to be calculated.

Radium A and radium C both emit α rays, whilst radium B and radium C emit β rays, and on this account the activity curves will be different according to whether the α rays or the β rays are employed for measuring the activity.

The α-ray activity M is given by the equation

$$M=\lambda_1P+k\lambda_3R, \quad . \quad . \quad . \quad (24)$$

where k is the ratio of the ionisation produced by an α particle from radium C to that produced by an α particle from radium A under similar conditions. The β-ray activity N is given by the equation

$$N=\lambda_2Q+l\lambda_3R, \quad . \quad . \quad . \quad (25)$$

where l is the ratio of the ionisation produced by a β particle from radium C to that produced

by a β particle from radium B. The theoretical curves calculated from these equations, when appropriate values are inserted for the ratios k and l, are shown in *Fig.* 17; curves

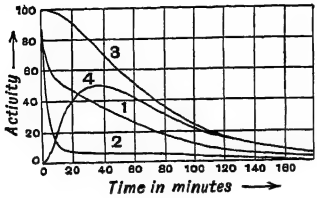


FIG. 17.

(1) and (2) show the variation of α-ray activity of radium active deposit after a long and a short exposure respectively to the active material; curves (3) and (4) the variation of the β-ray activity of radium C only after a long and a short exposure to the active material. The curves do not give the relative magnitudes of the ionisation currents produced by the α and the β rays respectively.

The above cases will suffice to show the usefulness of the theory to calculate the decay curves of radioactive products. A fuller account of the theory and its application will be found in Rutherford's *Radioactive Substances and their Radiations*.

§ (21) CHEMICAL CHARACTER OF THE RADIO-ELEMENTS.[1]—Each of the radio-elements appearing in Table I. has distinct chemical as well as radioactive characteristics, and although many of the products are exceedingly short-lived their complete chemical properties have been unravelled. In some instances the radioactive product may be prepared "chemically" pure, in which case its properties may be examined by chemical methods similar to those adopted in the case of ordinary elements —radium salts were prepared by Madame Curie sufficiently pure to carry out an investigation on the atomic weight of radium. In other cases a product may be obtained "radioactively" pure, in which case it is entirely free from other radioactive material, but is mixed with a certain amount of inactive matter. This may be the case when a radioactive substance is separated by precipitating it in the presence of an element resembling it in chemical properties. Lastly, a product may be prepared "radio-chemically" pure, in which the radio-element may be mixed with a certain amount of inactive matter, but is free from substances chemically analogous to itself. Such is the case with a product prepared by the method of recoil or as active deposit. The inactive material present in the last two cases will not interfere with the detection and measurement of the radioactive substance, but it will absorb the radiation emitted, a correction for which can be applied if necessary.

The results of the purely chemical study of the radio-elements has led to the division of the elements into two main classes:

(*a*) Those which possess distinct radioactive properties, but whose chemical characters are already completely known; and

(*b*) Those which, in addition to possessing a characteristic and distinct radioactivity, are new in chemical properties.

An instance of the first class is ionium. It resembles thorium more closely than any other member of Group III. of the Periodic Table and is found to be chemically identical in properties with it. The radioactivity of ionium is quite unique and different from any other radio-element, but in its chemical nature it is not new. All its chemical properties may be inferred from those of thorium, which can be obtained in large quantities, whereas ionium itself occurs in infinitesimal quantities, and even then it cannot be obtained "radio-chemically" pure, because thorium is present in all the minerals from which it may be obtained.

An instance of the second class is radium. It resembles barium very closely but not absolutely. In this case the chemical as well as the radioactive properties are unique, and we have a new type of chemical element.

§ (22) POSITION OF RADIO-ELEMENTS IN THE PERIODIC TABLE.—It was pointed out by Soddy that the expulsion of an α particle caused a change in the position of a radio-element in the periodic table by two places in the direction of diminishing mass. Fajans also drew attention to the fact that in a transformation involving the emission of β rays the resulting product is more electro-negative than the parent substance. Further knowledge of the chemical properties of the radio-elements has led to a complete generalisation as to the relation of the position of the radio-element in the periodic system and the nature of the change in which it is produced. The generalisation was independently arrived at by Russell,[2] Fajans,[3] and Soddy,[4] and may be stated thus:

In an α-ray transformation the product falls into a group two places lower than that to which the parent substance belongs, whereas in a β-ray transformation the product falls into a group one place higher than the parent substance.

The resultant grouping of the radio-elements is shown in Table V. It will be observed that the α-ray changes are far more numerous than the β-ray changes, and

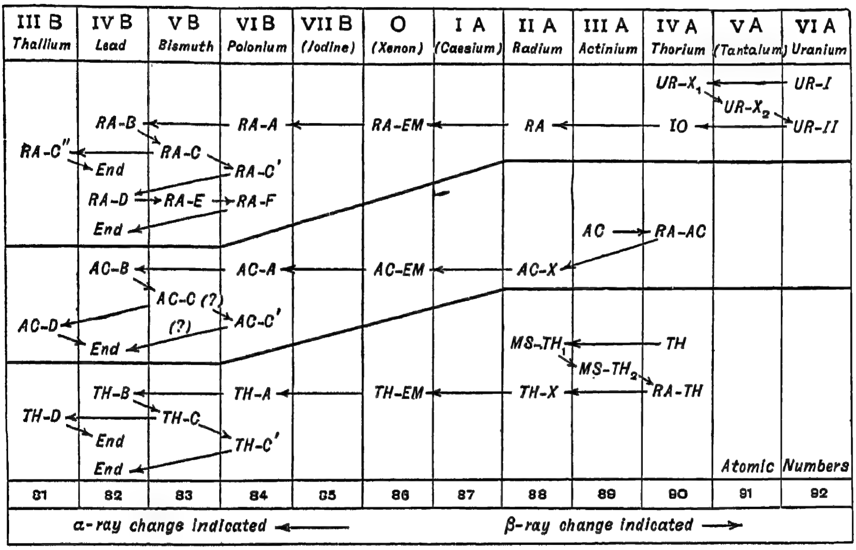
[1] See Soddy, *The Chemistry of the Radio-elements*, p. 43.

[2] Russell, *Chem. News*, 1913, cvii. 49.
[3] Fajans, *Phys. Zeits.*, 1913, xiv. 131 and 136.
[4] Soddy, *Chem. News*, 1913, cvii. 97.

in consequence there is a movement from the right to the left extending over the last twelve places in the periodic table from uranium to thallium. The members in the vertical columns of the table are non-separable from one another and from the title element, except when this is enclosed within brackets. Thus uranium X_1, ionium, radio-thorium, and radio-actinium are all chemically identical with thorium; they belong to the first class mentioned above, all their chemical properties being known "by proxy." To the second class of radio-elements belong, for example, mesothorium 1, thorium X, and actinium X, which are all chemically identical those elements; such tests would only distinguish between elements in different columns of the table, and consequently, would only pick out ten kinds of radio-elements, whereas radioactivity tests would show the existence of thirty-four, so that separate places in the periodic table do not correspond with single elements necessarily, but with single chemical types of elements.

When uranium I loses an α particle with the production of uranium X_1, each atom of uranium X_1 is deficient of two positive charges. Uranium X_1 changes to uranium X_2, and uranium X_2 to uranium II, with the loss of a β particle in each case, which brings the charge

TABLE V



with radium. Elements of the first class are similar to one or other of the last five elements in the periodic table, namely, uranium, thorium, bismuth, lead, and thallium. The members of the second class, which are new types of chemical elements, are identical with one or other of five new elements, namely, radium, polonium, actinium, the emanations, and "eka-tantalum." The chemistry of radium and the radium-emanation is completely known, and consequently all the other elements in the same column in the table are definitely known. Hence, apart from actinium, polonium, and "eka-tantalum," the chemistry of the whole of the radio-elements is known.

§ (23) ISOTOPES.—Elements with identical chemical properties are called "Isotopes," thus radium B, radium D, thorium B, and actinium B are all isotopes of lead. Chemical tests would not be able to distinguish between on the atom to the same value as it had before any transformation took place. Uranium I and uranium II are identical in chemical character, and the net charge on the nucleus of each of the atoms is the same, but owing to the loss of an α particle in the transformation the masses of the atoms differ by four units. Hence it would appear that it is the net charge on the nucleus of the atom and not its mass that determines its chemical properties.

Since the expulsion of an α particle displaces the element two places in one direction and the expulsion of a β particle displaces it one place in the opposite direction, each of the successive places in the periodic table corresponds with unit difference in the net positive charge on the nucleus of the atom. The magnitude of the positive charge is now known to be exactly equal to the number of the element in the periodic table when the

elements are arranged in order of atomic weight. This number is called the "atomic number" of the element. Thus the atomic number of hydrogen is 1, helium 2, etc., up to that of the last and heaviest atom of uranium with atomic number 92.

Not only do the isotopic elements possess the same net nuclear positive charge and the same number of electrons in their external systems, and are chemically identical and inseparable; their common purely physical characteristics, such as spectrum and volatility, are also found to be identical.

It has recently been found that this property of isotopism of the elements occurs frequently amongst the elements. Aston[1] has found that neon, for instance, is a mixture of two isotopic gases of atomic weight 20·00 and 22·00. Similarly chlorine contains no atoms of atomic weight 35·46, but consists of four isotopes with atomic weights 35·00, 36·00, 37·00, and 38·00 respectively. Several other elements have been examined, the conclusion arrived at being that all the atomic weights, except that of hydrogen, are exact integers, and that the fractional values found for some of the elements are due to a mixture of two or more isotopes.

§ (24) THE END PRODUCT.—The ultimate products of all the disintegration series in all branches end in the same place in the periodic table, namely, the place occupied by lead. The atomic weight of ordinary lead is 207·2. According to theory, lead produced from uranium should have an atomic weight 206, and that from thorium an atomic weight 208. Hence lead prepared from uranium minerals should have a lower atomic weight than that prepared from thorium minerals. Experiment[2] has shown this to be the case; the atomic weight of radio-lead from uranium minerals has been found to be 206·05, and that of radio-lead from thorium 207·9. Probably ordinary lead is a mixture of these two isotopes. It has also been found by experiment that the densities of the different kinds of lead are different just in proportion to the differences in their atomic weight, hence their atomic volumes must be the same, which is to be expected if isotopic atoms have identical shells of electrons but nuclei of different masses.

BIBLIOGRAPHY

E. Rutherford, *Radioactive Substances and their Radiations*, Cambridge University Press, 1913; W. Makower and H. Geiger, *Practical Measurements in Radioactivity*, Longmans, Green & Co., 1912; W. H. Bragg, *Studies in Radioactivity*, Macmillan & Co., 1912; F. Soddy, *The Chemistry of the Radio-Elements*, Part I., Longmans, Green & Co., 1917; *The Interpretation of Radium*, John Murray, 1920.

E. A. O.

[1] Aston, *Nature*, 1920, cv. 8 and 547, cvi. 468; *Phil. Mag.*, 1920, xxxix. 611, xl. 628.
[2] Soddy, *Nature*, 1917, xcviii. 469.

RADIO-ELEMENTS, CHEMICAL CHARACTER OF. See "Radioactivity," § (21).

RADIO-ELEMENTS, POSITION OF, IN THE PERIODIC TABLE. See "Radioactivity," § (22).

RADIOGRAPHY: X-ray photography. See "Radiology," § (29).

RADIOLOGY

I. INTRODUCTORY

§ (1) THE NATURE OF X-RAYS.—The study of the X-rays[3] now occupies such a prominent position in physics and medicine, and has led to such momentous results in a variety of directions, that it is difficult to realise that it is only 25 years ago that the rays were discovered by Professor Röntgen, and that it was only just prior to the war that a long controversy as to their nature was stilled.

The problem had attracted many minds, for the ability of the rays to pass through opaque bodies was wholly unprecedented. The explanation of the anomaly was obviously bound up with the nature of the rays, but, despite shrewd guesses, the secret was withheld from us for nearly 20 years.

We now know that the X-rays are another manifestation of radiant energy, of which light and heat are familiar examples. Indeed, the X-rays resemble light rays in almost every particular, the chief difference being that the X-rays have wave-lengths about 5000 times shorter. It was this very minuteness of wave-length—a distance of the same order as the sizes of atoms—that defeated all our earlier attempts to direct and sort out the rays. All our highest quality polished surfaces are inconceivably rough for such a purpose, and it was not until Nature herself was found to have provided an instrument of the requisite delicacy—in the shape of crystals which can function as diffraction gratings—that we began to analyse and sort out X-ray beams with much the same ease as in the case of visible light.

There are further parallelisms between X-rays and light rays. For example, we know that the spectrum of a hot body consists under suitable conditions of white light (which is a mixture of all wave-lengths), superposed on which are certain spectrum lines whose wave-lengths are characteristic of the radiating material, *e.g.* the D lines of sodium, the H and K lines of calcium. In just the same way an element when caused to emit X-rays not only gives out general radiation (which is a continuous spectrum of wave-lengths) but under suitable conditions impresses its own characteristic lines—K, L, M—on the general

[3] See also "X-Rays," Vol. II.

radiation. It should be added, however, that the X-ray spectrum of an element is much simpler than its light spectrum.

Just about a single octave of light waves are visible to the eye. Their spectroscopic examination has been conducted mainly with the diffraction grating, the distance between the rulings of which is comparable with the wave-lengths to be measured. With the help of special gratings and vacuum spectrometers, Schumann, Lyman and Millikan have extended the measurements some 4 octaves onwards into the ultra-violet. Then comes a gap of 4 octaves of rays in a region not yet explored, and finally some 7 or 8 octaves of X- and gamma-rays, of which the radiologist uses about 3 octaves (*Fig.* 1).

It is interesting to note that the gap between X-rays and ultra-violet rays has in a sense been already bridged, for Millikan has recently found the characteristic X-ray lines of carbon in the extreme ultra-violet region.

wave is very high and we get a high-frequency or "hard" X-ray. If the change of speed is less, the frequency is less and we get a "softer" or less penetrating ray. With much slower electrons, light rays may be similarly produced. Always, however, we find that the frequency of the wave is proportional[1] to the energy change of the electron. There will be a proportion of encounters where the whole of the energy is transferred, and in these cases the frequency will reach an upper limit. Below this limit we find every variety of energy-content depending on the energy of the electron involved.

The reverse effect is equally true. If X-rays (or light rays) strike a substance they may give up all their energy to moving electrons, or they may give up only a part, the rest being transferred to a series of groups of rays, all characteristic of the atom of the material. The energy balance-sheet can be fully set out, the several items all being definite and specific. The relation is not quite so simple as the

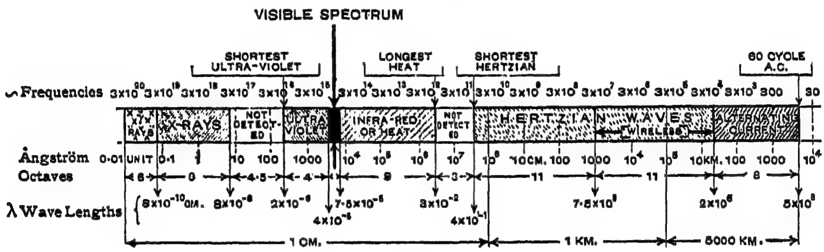


FIG. 1.

The study of this missing group of octaves bristles with interest and difficulties. It would seem as if a big absorption band were included; at either end of the gap the vacuum spectrometer is necessary. Moreover, it almost looks as if the grating method would fail us over this region. The wave-lengths are too small for our artificial gratings and too big for crystal gratings. We may have to wait for a new weapon of attack.

Following is a short table of some of the wave-lengths, in Ångström units, *i.e.* 10^{-8} cm.:

Visible light	7200–4000
Ultra-violet light	4000–200
X-rays	12 to 0·06
γ-rays	1·4 to 0·01

§ (2) X-RAYS AND ELECTRONS.—Experiment has shown the most intimate relationship between the X-ray and the electron—either is the manifestation of the other. The electron, the unit of negative electricity, is the active manifestation of the X-ray—innocuous in itself. Whenever an electron has its speed suddenly altered, an electromagnetic wave is produced. If the alteration of speed of the electron is very great the frequency of the

general case, but the exchange and partition of energy are equally precise.

The reversible process we have just described is of universal application in Nature. There is, for example, little doubt that the X-rays play a prominent part in atmospheric electricity. The earth is not an electrically neutral body, but its surface may be considered to be covered with a layer of negative electricity, and this gives rise to an electrical field in the atmosphere. The rate of alteration of potential is found to decrease with the altitude; the potential gradient being about 150 volts per metre on the ground and only about 2 volts per metre at a height of 9 kilometres—as we know by balloon tests. In other words, the atmospheric conductivity steadily increases the higher we go, and the rapidity of the increase suggests very large values at greater heights.

Some of this conductivity, we know, is due to radio-active emanations from the soil, but we are led to infer from the increase of conductivity with height that the majority is due to some agent external to our globe. Modern opinion favours the view that the effect is produced by very high-speed electrons

[1] See "Quantum Theory," § (3).

ejected from the sun with very nearly the speed of light itself. Some of these strike the atoms of the outer atmosphere, very penetrating X-rays are generated, and thus the whole depth of the atmosphere may be permeated by these electrons through the intermediary of their more penetrating alias the X-rays. Thus the earth's negative charge which is being continually dissipated by the action of the potential gradient in the atmosphere is as steadily replenished by a current of electrons passing downwards. It may be added that the conductivity of the air diminishes at night and during a solar eclipse.

One is tempted also to believe that in view of the temperature and gigantic electrical disturbances in the sun—as Hale's work has shown and Eddington's speculations would indicate—there may be an emission of X-rays from the sun itself.

One other source of X-rays in Nature may be referred to—the γ radiation of the radioactive elements. It will suffice to say that while some of the γ-rays can be exactly imitated, others are much more penetrating than any X-ray we have been able to generate artificially.

Although the electron is ubiquitous, it escaped detection until Crookes conducted his famous experiments in discharge tubes at low pressures, and so reduced the number of molecules present that instead of the electron being absorbed and suppressed within a mm. or so, as it would be at atmospheric pressure, it could now travel great distances without encountering more than say, 100 or so atoms, the majority of which it passed clean through without being deviated in any way. The high speed it received from the potential in the discharge tube gave it, so to speak, an innings.

We have anticipated matters somewhat, but it will now be convenient to review the main features of a discharge tube and their historical development.

§ (3) The Phenomena of a Discharge Tube.—When a high-potential discharge[1] is passed between two electrodes within a glass tube, the gas pressure in which is gradually reduced, the tube displays a succession of appearances which vary with the pressure. At atmospheric pressure the discharge takes the form of a spark which, as the pressure is lowered, loses its noisy character and is replaced by a bundle of irregular streamers, which, after a time, broaden out and fill the bulk of the tube with a pink glow, known as the positive column.

At a pressure of a few millimetres of mercury the cathode (the electrode by which the current leaves the tube) slowly becomes covered with a luminous glow—the negative glow—while between the positive column and negative glow comes a darker region, diffuse in outline, called the Faraday dark-space. All this time the rarefied air is increasing in conductivity, as evidenced by the shortening of the alternative spark gap.

As the exhaustion proceeds the positive column may break up into thin fluctuating striations, which presently widen and diminish in number. The negative glow proceeds to detach itself from the cathode, while a new film forms and spreads over the surface of the cathode. The two parts of the negative glow are separated by a dark region called the Crookes or cathode dark-space, the outline of which is sharply defined and runs parallel with that of the cathode. As the pressure is reduced the dark-space increases in size, its thickness often being used as a rough measure of the pressure.

At still higher rarefactions both positive and negative glows become darker and more indefinite, and the cathode dark-space grows until finally its boundaries touch the glass walls of the tube, which then begin to fluoresce.

Meanwhile the conductivity of the rarefied gas has been steadily lessening until, finally, it is possible to get the exhaustion so complete that no discharge can pass.

§ (4) Cathode Rays.—The study of the fluorescence of the glass wall was initiated by Plücker in 1859. He was followed in the 'seventies by Hittorf and Goldstein in Germany, Puluj in Austria, and Crookes in England. It was soon ascertained that the fluorescence was produced by some form of radiation (called Kathodenstrahlen by Goldstein) which sets off from the cathode and travels in straight lines. For nearly thirty years the English and German schools of physics disagreed as to the nature of the cathode rays, but Sir J. J. Thomson in 1897–98, in a series of classical experiments, was able to show that the cathode rays were bodies of sub-atomic size, moving with prodigious velocities, velocities which are comparable with the speed of light. Johnstone Stoney had previously suggested the name electron for the unit of electricity, and the suitability of the expression for the cathode rays was at once recognised. Thomson showed that each cathode ray had a mass about 1/1800 of that of the hydrogen atom and carried a negative electrical charge, agreeing in amount with that carried by the hydrogen ion in liquid electrolysis.

The bulk of the great energy of the cathode rays is dissipated as heat when the rays strike an obstacle. If the rays are concentrated by using a concave cathode or other focussing device, enormous heat can be generated with heavy discharges. Cathode-ray furnaces, working on this principle, have been con-

[1] See also "Electrons and the Discharge Tube," Vol. II.

structed, in which any known metal can be melted or vaporised.

By reason of their electric charge cathode rays are readily bent under the action of electric or magnetic force, and this property of the rays has found practical application in Braun tubes and other forms of oscillograph. The cathode rays, having no inertia, are able to follow the most rapid vagaries of an electric or magnetic field.

§ (5) Positive Rays.—In 1886, Goldstein, using a perforated cathode, noticed that a stream of rays emerged from the tube in a direction opposite to that of the cathode rays. Wien later observed that these "Kanalstrahlen" carried a positive charge. The deflection in electric and magnetic fields showed that the positive rays[1] (as they were called later by Sir J. J. Thomson) were molecular in size. J. J. Thomson and Aston have done extended work on the subject and derived highly important results. Positive rays play an important part in gas X-ray tubes, as we shall see presently.

§ (6) Discovery of X-rays — Röntgen Rays.—It was in the autumn of 1895 that Professor W. K. Röntgen at Würzburg, Bavaria, discovered the rays which now bear his name. During the course of a research on invisible light rays, he had enclosed a discharge tube within a screen of stout black paper. On passing the discharge he noticed that a fluorescent screen, lying on a table some distance away, shone out brightly. By interposing obstacles Röntgen traced back the unknown or "X" rays to their source, which proved to be the region of impact of the cathode rays on the walls of the tube. The feature of the new rays was their uncanny ability to penetrate many substances quite opaque to light. The degree of penetration was found to depend roughly on the density. For example, flesh is more transparent than bone, and accordingly the bones stand out dark in the shadow caused on a fluorescent screen. Röntgen saw at once the immense importance of his discovery to surgery, and communicated his results to the Physico-Medical Society of Würzburg in November 1895.

The main features of the rays were soon discovered by an army of workers. It was found, for example, that X-rays travel in straight lines, cannot apparently be refracted or reflected, do not carry an electric charge, and possess the property of ionising or imparting temporary electric conductivity to a gas.

§ (7) Detecting the Rays.—Although X-rays are not in the visible spectrum they can be detected photographically or by their power of exciting fluorescence in screens made of salts such as barium platino-cyanide, or calcium tungstate. Another method makes use of the ionising ability of the rays and measures the conductivity so produced in a gas. Certain chemical reactions are also induced by the rays and can be made to serve as the basis of various methods of detection and measurement.

[1] See "Positive Rays," Vol. II.

X-rays can penetrate all substances to a greater or less degree, and in general the shorter the wave-length the higher is the penetrating power. The penetrability of a material by a given beam of rays is governed by the number and mass of the atoms it encounters, that is, by the atomic weight and thickness. Chemical combination or temperature is without effect on the absorbing power of an atom. We have already mentioned that the rays travel in straight lines, and thus it will be seen that an X-ray photograph or radiograph is essentially nothing but a shadowgraph. Radiography was the first and still remains the most important application of the rays, and in the hands of the medical man has found enormous application. The late war brought this home in unexampled fashion, and the services which radiology then rendered can scarcely be overestimated.

The subject is referred to later, but mention may here be made of the necessity of protecting the X-ray operator from the rays. As many of the early workers discovered to their cost, indiscriminate exposure results in dermatitis, which may be followed by dangerous cancerous growths. With hard rays derangements of the internal organs and impoverishment of the blood corpuscles may result. Nowadays every precaution is taken and such casualties rarely occur. Heavy lead screening in some form limits the beam of rays and protects the worker.

II. X-ray Tubes

§ (8) Early Forms.—The vacuum tube with which Röntgen made his discovery had a flat cathode, the cathode rays impinging on the glass walls. Experience soon showed the way to improvements. Campbell-Swinton inserted a platinum target obliquely in the path of the rays, and later Sir Herbert (then Professor) Jackson replaced the flat cathode by a concave one, so that the cathode rays were brought to a focus on the target. The result was that exposures were enormously shortened and, owing to the small area of emission of the X-rays, the resulting photographs were improved out of all recognition in definition and detail.

The present-day method is essentially unaltered. The electrons (cathode rays) are given enormous speeds (of the order of 50 to 100 thousand miles a second) by means of high voltages, and are directed on a heavy metal anticathode or target. As a producer of X-rays the arrangement is still extremely

inefficient, although we take steps to increase the chances of an effective collision by choosing a target of high atomic weight or number.

As already mentioned, almost all the energy of the electrons is degraded into heat, and for this reason it is essential that the target shall be of a very refractory metal. Tungsten (with a melting-point of over 3000° C.) is nowadays almost always employed for a target, though platinum and other metals find application for certain purposes.

The X-rays radiate uniformly in all directions from the focus, travelling in straight lines just as light rays radiate from a lamp. The X-ray bulb is comparable to an X-ray lamp, in which the voltage applied to the bulb corresponds to the temperature of a luminous lamp. If we raise the temperature of the latter we increase the intensity and at the same time shorten the average wave-length; so with the X-ray bulb, if we raise the voltage we increase the intensity and shorten the average wave-length. In practice the voltages employed range up to 200,000 or even more. The quantity of radiation is dependent on the current through the tube, and a milliampere is a convenient measuring unit for the purpose.

There are two main types of X-ray bulbs in use (*a*) the hot cathode tube, (*b*) the gas tube. The Coolidge tube, invented by Dr. Coolidge in America, is the chief representative of the first class, in which the electrons are produced from a cathode consisting of a spiral of tungsten wire, raised to a white heat by an electric current. The vacuum in the tube is very high, and no discharge can pass if the cathode is not heated. The Coolidge tube has the valuable property of precise and reproducible control with a great range, advantages which cannot be claimed for the gas tube. In the gas tube very complete exhaustion is not attempted; a trace of residual gas is deliberately left in the tube, and this serves as a constant source of electrons through shock ionisation.

If we compare the characteristic curves of these two types of X-ray lamps by plotting current against voltage, we find differences which are fundamental (*Fig.* 2). Under the conditions in which a gas tube operates the current increases steadily with the voltage, while in the case of the Coolidge tube the current is independent of the voltage. In the latter case the current is limited only by the number of electrons emitted, which number increases or decreases with the temperature of the cathode filament. Thus we can alter either voltage or current independently of each other, and this fact gives the hot cathode tube a great advantage over the gas tube, in which independent control of voltage and current is impossible.

The hot cathode tube thus utilises its "saturation" current, and for that reason is much less affected by changes in the wave form of the exciting potential than is a gas

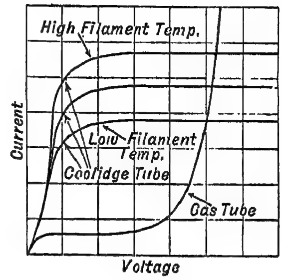


FIG. 2.

tube. On the other hand, the absence of saturation in a gas tube leads to a more effective use of very high voltages, and it is found that, at any rate, at low gas pressures a gas tube gives about twice the X-ray output of a Coolidge tube for the same milliamperage and voltage (*Fig.* 3, Dauvillier). But the gas tube is far from being the equal of the Coolidge tube as regards control and reliability, though experience counts for a good deal.

It may be added that both types of tube

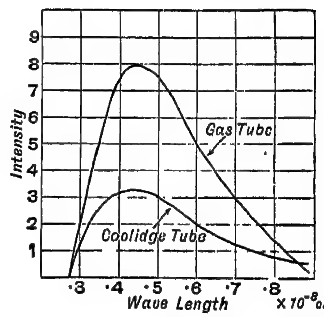


FIG. 3.

give heterogeneous X-rays, and there is little to choose between them in this respect.

§ (9) THE ELECTRODES OF AN X-RAY TUBE.—*The cathode* of the present-day X-ray gas tube is made of aluminium of robust design. It is mounted facing the anticathode or target, the distance between the two being a matter of nicety for the maker, who has to be guided largely by his experience and knowledge of the work for which the tube is intended. The lower the pressure or the "harder" the tube, the farther the focus of the cathode rays recedes from the cathode and the smaller the focal spot.

The material and size of the cathode are both important. Aluminium possesses advantages because it displays but little "sputtering"

or blackening, a feature which is common to almost all metals when functioning as a negative electrode in a vacuum. Aluminium shows the effect but slightly, but with very heavy discharges the central portion of the cathode may be melted, the molten globules so formed being projected right across the tube in some cases. No particular harm is usually done, except that the curvature of the cathode may be changed and the focal spot of the cathode rays moved in consequence.

The choice of metal for cathode is also important from the point of view of the occluded gas. There is little doubt that in the case of a gas tube the gases contained within the metal of the cathode play an important part. This is probably the explanation why some experimenters have found it useful to keep the cathode cool by some device.

The cathode of a Coolidge tube is described later.

§ (10) The Anticathode.—The primary essentials in an anticathode, whether for gas or hot-cathode tube, are:

1. A high atomic number to secure a large output of X-rays.
2. A high melting-point and a high thermal capacity to prevent the target fusing under a heavy discharge.
3. A high thermal conductivity to assist in dissipating the waste heat.
4. A low vapour pressure to avoid distillation and condensation of the metal on the walls of the tube.

In the light of the above requirements, tungsten stands almost alone as a material for anticathodes. Platinum was at one time almost exclusively employed, but tungsten has a much higher melting-point (3200° C. as against 1755° C.), has twice the thermal conductivity, and but a slightly inferior radiation value.

The very large heating effects experienced by the anticathode of a modern X-ray bulb have provided scope for much ingenuity in getting rid of the surplus heat. The anticathode is in many tubes kept cool by means of a water reservoir or a stream of air. In other cases the massiveness of the anticathode is increased by surrounding the tungsten target with copper, the support of which extends to the outside of the tube and is there provided with radiating fins, incidentally necessitating some very fine glass manipulation.

§ (11) The Anode.—The modern X-ray gas bulb is almost always provided with an additional anode of aluminium, which is connected externally with the anticathode. The precise benefits of the separate anode are distinctly doubtful, and in tubes of the Coolidge type it finds no place.

§ (12) The Gas Tube.—The gas tube depends for its action on the presence of a few ions in the residual gas in the tube. These ions or electrified atoms have their velocities increased by the electric field, positive ions being drawn to the cathode and negative to the anode. The positive ions bombarding the cathode release electrons in abundance which, being attracted to the anode, ionise freely by shock or collision those atoms encountered *en route*, generating more positive ions and more electrons. The electrons which hit the target generate X-rays and the cycle of operations continues so long as the voltage is applied.

The positive ions or positive atoms thus play a fundamental and essential part in the ionics of a gas tube. They are also responsible for one or two other effects, the elucidation of which has been very puzzling. One of the great difficulties in exact work with the gas tube is the continual tendency of the gas pressure to change. One would first look to the electrodes which, depending on the conditions, may either emit or absorb gas and do so control very materially the well-known "crankiness" of a gas tube. But it is found that, provided the current is not too heavy to overheat the electrodes, there is a continual and apparently unlimited disappearance of gas, more especially at high voltages, and ultimately the vacuum becomes so high as to render the tube unusable. To cut a long story short we now know that some of the positively charged atoms of gas by reason of their high velocity (about 500 miles a second) actually crash into the glass walls of the tube and are mechanically trapped there, an effect which is enhanced by the presence of volatilised metal.

Many devices have been introduced from time to time to overcome this hardening effect. In some cases, when the tube becomes too hard, the discharge is caused to pass through a small annexe containing absorbent material, such as asbestos, which liberates enough gas to soften the tube. Another method commonly employed is to rely on the diffusion of gas through a small platinum tube, which can be heated by a small flame.

If a gas tube is overloaded the result is sometimes to harden the tube, sometimes to soften it, depending largely on the behaviour of the cathode. Further, if the discharge is sufficient to make the target red-hot, it may then give off so much gas as to necessitate re-exhaustion of the bulb.

A very commonly met source of failure of gas tubes is due to cracking of the glass, almost always in the region round the cathode. In many cases this is preceded by a roughening of the glass at this point. These effects are produced by positive rays striking the glass walls. The positive rays are probably also responsible for the melting of the aluminium

cathode under very heavy discharges. There is little doubt that the positive ions produce the luminous blue glow which often accompanies high-voltage discharges, and also the positive electrification which is present on the glass walls of the bulb of a gas X-ray tube. Positive ion effects are always found to persist so long as the electrodes are giving off gas.

A common defect with gas tubes is the wandering of the focal spot. Such movement is, of course, prejudicial to good definition in radiography. Further, it is not unusual, after the tube has been subjected to a heavy discharge, to find that the focus has permanently changed its position.

Another characteristic of a gas tube is that the breakdown voltage is a good deal higher than the running voltage. This is due to the fact that in the gas tube the number of ions initially present is very small. When the circuit is closed the number of ions produced by collision increases enormously, and the voltage across the terminals of the tube falls in consequence.

The gas tube received a good deal of attention in Germany during the war. In the so-called "boiling tube" of Muller, the gas tube is employed to generate very penetrating X-rays primarily for use in deep therapy. The gas pressure is not only low, but tending to get lower. To prevent emission of gas from the (platinum) anticathode, it is kept at a constant temperature by boiling water, and water-cooling is also adopted for the cathode. The focus is very broad. The tube is operated at close on 200,000 volts and with small currents, 2 to 3 milliamperes, a condition which assists the hardening tendency. Puffs of gas are introduced by an osmosis tube heated by a small flame ignited and operated through a relay by an automatic regulator which is controlled by a milliammeter in series with the tube. In action the water quietly boils, and the tube may be run for hours at a time at a constant milliamperage. It is claimed that there is a greater proportion of homogeneous end radiation when the X-rays are filtered than there is from a Coolidge tube working under the same conditions.

Whether that is so or not, it must not be forgotten that an increase of potential always tends to render a beam of X-rays more homogeneous. Further, the spark-gap readings on a gas tube tend to indicate voltages higher than those which are actually operating the tube.

§ (13) THE COOLIDGE TUBE.—Consideration of the properties of a gas X-ray tube indicates that most of the limitations are incidental to the presence of positive rays. Dr. W. D. Coolidge[1] of the Research Laboratory of the General Electric Company, Schenectady, New York, saw that most of these disabilities would disappear if the vacuum in the tube were high enough, so that the positive ions did not play an essential rôle. This necessitated generating cathode rays by some other means, and Coolidge turned to the work of Richardson,[2] who had made quantitative measurement of the electrons produced by heating a negatively charged metal. There was considerable divergence of opinion at the time as to the mechanism of the effect, and certain workers had suggested that in the absence of gas such thermionic discharges would cease altogether. Langmuir[3] in some experiments on the thermionic currents in the case of tungsten, showed that the effect was a real one, and that instead of dying away the currents actually increased, up to a certain limiting value, as the tube became freer and freer from gas. In brief, Langmuir showed that a hot tungsten cathode, in a vacuum as high as it is humanly possible to make it, emits electrons continuously at a rate determined only by the temperature.

Dr. Coolidge was thus led to evolve the Coolidge X-ray tube, the main features of which are as follows:

The cathode consists of a small flat tungsten spiral which can be electrically heated by an independent circuit. Surrounding the spiral is a tube or bowl of molybdenum, which serves to focus the cathode ray stream upon the target. The temperature of the filament varies from about 1600° to 2300° C. The bowl is connected electrically to the cathode, and the anticathode or target functions also as anode. In the earlier or "universal" type of tube the target consists of a block of tungsten with a molybdenum stem. This design has been modified in later types.

Unusual precautions are taken to render the exhaustion of the bulb as complete as possible. Langmuir pumps with liquid air traps are used, and the whole tube, while connected to the pump, is heated in an electric oven to just short of the softening point of glass (about 470° C.). Periodically the tube is operated with heavy discharge currents, so that the tungsten target is raised to a white heat. The test of adequate exhaustion is the complete absence of a luminous discharge and phosphorescence of the glass walls. It is found that the time of exhaustion can be considerably shortened by previously heating the electrodes to a high temperature in a vacuum furnace. Metal thus treated may be left about in the air for some days and only absorb but a small fraction of the gas originally contained in it.

The great merit of the Coolidge tube is its capacity to reproduce completely any specified

[1] *Phys. Rev.*, 1913, ii. 409.

[2] *Roy. Soc. Proc.*, 1903, lxxi. 415.

[3] *Phys. Rev.*, 1913, ii. 450.

set of conditions. The output (or quantity) of X-rays depends only on the temperature of the filament. The penetrating power (which depends upon the wave-length) is controlled only by the voltage across the terminals of the tube. Furthermore, the tube allows current to pass in only one direction, so long as the conditions are such that the target is not raised in temperature to more than a dull red heat. The tube is, in fact, capable of rectifying its own current, when supplied from an alternating source. The latest form of radiator tubes permits the passage of a steady current of as much as 30 milliamperes. Care should be taken that the tube is not overloaded, as otherwise, if the target attains a white heat, a considerable amount of inverse current will be allowed to pass, with detrimental and ultimately disastrous effects on the bulb.

Another feature of the Coolidge tube is that the starting and running voltages are the same. As the positive ions play no appreciable part, there is little or no consequential heating of the cathode, and no evidence of appreciable cathodic sputtering. The Coolidge tube in operation shows none of the fluorescence of the glass displayed by gas bulbs, a result due to the fact that the walls of the tube become negatively charged, and so, unlike an ordinary gas bulb, repel the "reflected" cathode rays from the target.

In the prolonged use of a Coolidge tube a small amount of gas is usually liberated, as is indicated by a small drop of the milliamperes through the tube, which can be corrected by raising the filament temperature; such gas is promptly re-absorbed when the discharge is stopped.

Coolidge, by the use of the pin-hole camera, has shown that, in addition to the main body of X-rays from the focus, there is a very considerable emission of extraneous X-rays from the entire surface of the anticathode, which must therefore be bombarded by cathode rays from some source or other. We must look to the "reflected" rays for an explanation. These rays, which leave the focus with a velocity almost as high as that with which they approached it, are repelled by the cathode and the glass bulb, both negatively charged, and are compelled to return and collide once more with the anticathode. Doubtless in many instances this procedure occurs again and again, X-rays being produced at every collision. The pin-hole camera shows that, except in the region of the focus, the intensity of the X-rays is much the same, both back and front of the anticathode. Measurements show that the rays, whether from back or front, differ but little in penetration. Coolidge has devised a variety of apparatus for reducing the magnitude of the stray radiation, but in the end came to the view that the extent of the defect did not justify the loss of simplicity involved in providing a remedy. The use of a lead diaphragm as small as possible and as near the tube as possible is the best cure.

Coolidge has also developed various modifications of the original design of tube. In the radiator type of tube, referred to above, the anticathode is made up of a solid bar of copper, which is brought out through the glass to a copper radiator (*Fig.* 4). The head of the anticathode consists of a mass of specially purified copper, which is first cast in a vacuum around a tungsten button, and is then electrically welded to the stem. In this type of tube the great bulk of the heat generated is conducted away by the radiator, and, as a result, it has become possible to make the glass bulb very small, a little under 4 in. in diameter. It may be added that the type of anticathode just described greatly increases the difficulty of exhaustion. In a still more recent model the glass walls have been made very thick, about a quarter of an inch. One effect of this is, for some curious reason, greatly to increase the steadiness of the discharge. Lead glass is used with a thin transparent window, and the tube thus affords protection for the operator. It has been found possible to manufacture a workable lead glass which affords as much protection as one-quarter the same thickness of sheet lead.

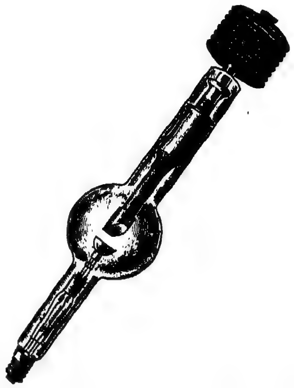


FIG. 4.—Radiator Type of Coolidge Tube.

In some experimental models Coolidge has worked with anticathodes cooled by currents of water. Such tubes permit enormous X-ray outputs. One tube ran continuously for 68 hours at 100 milliamperes and 70,000 volts. Other tubes have been run continuously for many hours at 200 milliamperes, the power input being of the order of 14 kilowatts. Doubtless this figure will be increased as time goes on.

The G.E.C. of America has adopted the methods of mass production to the various

forms of Coolidge tube, and is now turning out over 100 tubes a day. The bulbs and glass parts are blown in moulds at the glass factory, and the operation of assembly is carried out by girls with the aid of glass-blowing machines.

There is one feature of the Coolidge tube to which reference may be made. As already remarked, in consequence of the low pressure, "shock" ionisation of the residual gas is negligible or practically so, and the work of carrying the current is left solely to the electrons. Thus the space between the electrodes is filled with carriers of one sign, with the result that at high current densities there is an appreciable obstructing effect due to electrostatic repulsion between the electrons crossing over and those following. This "space-charge" sets an upper limit to the current through a Coolidge tube at high filament temperatures. Thus the current through a Coolidge tube may be set an upper limit either by the filament temperature or by the space-charge. The restricting effect of the space-charge can be lessened by raising the voltage or by introducing positive ions in some fashion, *e.g.* by a trace of gas. In the case of very heavy momentary discharges tungsten vapour is produced at the focal spot, and this also serves greatly to diminish the tube resistance.

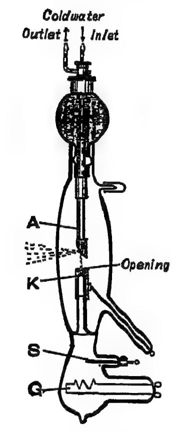


FIG. 5.

§ (14) LILIENFELD TUBE.—The Lilienfeld tube, introduced in 1913, and since extensively modified, may be said to act as a combination of hot cathode tube and gas tube, and, incidentally, is claimed to possess the advantages of both. In an annexe to the main discharge tube a hot cathode is separately excited by a moderate potential. The electrons pass through a hole in the main cathode and are there subjected to a much higher potential difference before they strike the anticathode (*Fig.* 5). Lilienfeld lays stress on the importance of using a coil discharge (as distinct from a transformer discharge) in that it yields higher momentary current densities. At high voltages (above 120 kv. and up to 170 kv.) the rays after filtering through 3 mm. of aluminium are stated to be homogeneous.

§ (15) METAL X-RAY TUBES.—From time to time various experimenters have worked with metal bulbs in an attempt to get rid of some of the energy limitations imposed by glass. Sir Oliver Lodge designed such a bulb in 1897, and since then Coolidge (*Fig.* 6), Siegbahn, and others have made use of them, and it is not unlikely that future commercial developments will be on such lines. Silica bulbs have also been used by Coolidge and other workers.

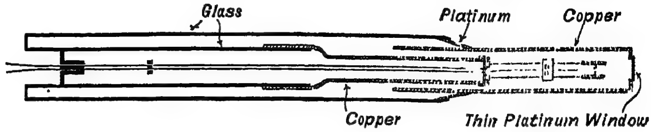


FIG. 6.

III. THE HIGH-POTENTIAL GENERATOR

The voltages which obtain in practice for exciting X-ray bulbs are, roughly speaking, of the order of up to 100,000 volts for radiography and superficial therapy, and up to 200,000 volts or more for deep therapy and metal-radiography.

§ (16) TRANSFORMERS AND INDUCTION COILS.—The high-potential generator is almost always a step-up transformer.[1] Less commonly, especially in this country, influence or static machines are employed, though they would have many advantages if they could be sufficiently improved to withstand atmospheric humidity.

It is customary in radiology to speak of "induction coils" and "transformers," though both are varieties of step-up, static, high-tension transformers. By an induction coil is meant an open-cored transformer which depends for its action on the interruptions of the primary current by an independent break or interrupter. By a transformer is implied a closed-core transformer fed with alternating current (almost always single phase) either straight from the main or (in the case of a D.C. supply) from the alternating side of a rotary converter. Such a transformer may be either oil-immersed or have "dry" insulation.

The coil ordinarily yields a "peaky" potential wave as compared with the approximately sinusoidal wave form of the transformer. With either type some sort of valve or mechanical rectifier is employed either to cut out or invert the half of the high-potential wave which would tend to pass in the wrong direction through the X-ray bulb. In the case of a coil and rotary interrupter the mechanical rectifier is a commutator mounted on an extension of the spindle of the interrupter. In the case of a transformer a similar

[1] See "Transformers, Static," Vol. II.

commutator is mounted, either on the shaft of a synchronous motor (if A.C. supply is used) or on an extension of the shaft of the rotary converter, as the case may be. If a valve is used instead, there are several types available, but choice is practically restricted to the hot cathode type such as the Kenetron. As already remarked, Coolidge tubes of the radiator type are self-rectifying.

The initial cost of a transformer outfit is approximately twice that of an induction coil outfit of corresponding power. A transformer outfit is rather more bulky, and any repairs are also usually more expensive. On the other hand, the efficiency of a transformer is roughly twice that of a coil (including a break and rectifier). Further, owing to the occasional vagaries of all interrupters, control is more precise and measurements are more definite with the transformer, and if A.C. is available and we dispense with the mechanical rectifier, there are no moving parts. Transformers produce greater heating effect on the target of the X-ray bulb, but this objection is met by arranging the rotating commutator so that it picks off only the regions round the crests of the loops, and thereby eliminates the less efficient lower voltages.

The induction coil is an empirically designed instrument, the present-day type of which is not fundamentally very different from the early models of Spottiswoode, although in detail it differs very considerably. The exact measurement of the performance of coils is difficult, and, as a consequence, coil makers have been led to adopt certain arbitrary standards of design which are based chiefly on practical experience. Some of the factors which arise in the design are conflicting, and it is in the methods of reconciling necessarily antagonistic factors that the skill of the coil designer finds chief scope.

The subject of induction coil design for X-ray purposes is a large one, and only certain broad conclusions can be touched upon here. A transformation ratio of from 50 to 200 and an efficiency of 0·3 to 0·6 are usual figures. Some form of sectional winding is adopted for the secondary coil, allowing about 4000 volts for every 1000 turns, and arranging that the outside diameter does not exceed 2½ times the bore. The resistance and, more especially, the self-induction of the secondary must be kept down. The primary should be capable of being connected directly to the 200-volt mains. The capacity of the condenser should be no greater than will prevent undue arcing in the interrupter. The interrupter should run at as high a speed as is expedient, and be of adequate design and robust construction.

About 15 lbs. of iron core should be allowed for every kilovolt-ampere input. The core may well have a length in the neighbourhood of up to 10 times the diameter. The primary windings should extend over almost the whole length of the core; the secondary windings over not more than the middle three-quarters, though care must be taken that this (the length of the secondary) is at least $\frac{4}{5}$ times the maximum spark length.

The induction coil is essentially a shock apparatus, and the shock excitation method of interruption may result in the presence of many superposed harmonics in the oscillation waves. These harmonics, which have high frequencies (several thousands a second), are reflected in the secondary circuit, where, from a practical point of view, they evince themselves in the reluctance with which they pass through an X-ray bulb. The resulting tendency to spark across the surface of the tube can only be met by lowering the gas pressure, by immersing the tube in oil, by lengthening the arms, or, of course, by suppressing the high-frequency waves before they reach the tube.

This is done in the so-called "symmetrical coil" which has been developed in Germany. In this apparatus two separate coils, mounted vertically side by side, have their secondaries connected in series, and also their primaries. In the two connecting leads between the secondaries are inserted a gas X-ray bulb in the one and an enclosed rectifying spark gap in the other. On each side of the spark gap and in series with it is a high resistance (water). The self-induction of the secondary circuit is low, but the resistance is very high and serves to damp out the high-frequency oscillations. The spark gap helps to enhance the breakdown potential of the gas bulb. High voltage (200,000) and low current (2 to 3 m.a.) are aimed at. An annular air space between primary and secondary assists natural cooling. A mercury break ($\sim = 47$) is used.

Interrupters.—Much of the progress that has been made with the performance of coils has resulted from the proper selection of interrupter. The hammer break, the accompaniment of most of the earlier coils, is now rarely fitted. The majority of present-day interrupters are of the motor-driven type, which employ mercury in a dielectric either of coal gas or a liquid such as paraffin oil. A large proportion of mercury interrupters are of the turbine variety, in which a jet of mercury is pumped against a series of rapidly revolving vanes.

The electrolytic or Wehnelt interrupter still finds favour with some workers. In the usual form it consists of two electrodes immersed in dilute sulphuric acid. The cathode is a large lead plate, the anode consisting of one or more platinum points, the exposed amount of which may be controlled by an adjustable porcelain sleeve.

There is still scope for much work on the design of interrupters, which may fairly be said to be the most untrustworthy feature of a present-day coil outfit. A large amount of energy is wasted in the interrupter, especially with heavy currents.

IV. The Measurement of X-rays

The output from an X-ray bulb must be specified with respect to (1) mean wave-length, quality, or hardness, and (2) intensity, *i.e.* quantity of rays per unit area.

The problem is complicated somewhat by the existence of two distinct types of radiation—(1) a general spectrum of X-rays with a large range of wave-lengths; (2) the "characteristic" or "monochromatic" rays which are wholly characteristic of the metal of the anticathode.

The proportions of these two classes depend on the conditions of discharge and on the material of the target. The characteristic radiations only appear when the exciting cathode rays are sufficiently fast. There is, in fact, a critical voltage for each metal, which is required in order to excite the characteristic rays, and the proportion of these rays increases rapidly as the voltage is raised above this critical limit.

§ (17) Methods of measuring Quality or Hardness.—The range of qualities of X-rays is very wide, embracing several octaves.

(i.) *Wave-length.*—We now know that the hardness, or penetrating power, of an X-ray is precisely defined by its wave-length—the shorter the wave-length, the harder the ray. The most precise means of measuring the quality of X-rays is by the crystal spectrometer. Measurements show that the majority of the wave-lengths of the X-rays, so far examined, lie between 10^{-7} and 10^{-9} cm.

The subject is dealt with elsewhere, but it has been shown by the Braggs, Moseley, and others that measurements of the diffraction of X-rays by crystals can be made to yield the wave-length of X-rays as well as the dimensions of the lattice-constant of the crystal concerned.

It may briefly be mentioned that in any crystal the atoms are regularly disposed in a network of intercrossing groups of planes, each of the planes in a group being parallel to and equidistant from its like neighbouring planes. The lattice-constant of a crystal is the distance separating the main atomic planes parallel to some specified crystal face.

In view of their importance in X-ray measurement we give below a table of X-ray wave-lengths for the three main series (K, L, M) of characteristic lines. Preceding this is a table of lattice-constants for several crystals on which extended measurements have been made.

(ii.) *X-ray Spectra.*—Up to now, about 16

Lattice-constants of Crystals

Crystal.	Lattice-constant.	Observer.
	$\times 10^{-8}$ cm.	
Rock salt, NaCl . . .	2·8140	W. L. Bragg, *Roy. Soc. Proc.*, 1913.
Calcite (cleavage face), $CaCO_3$	3·0290	Siegbahn, *Phil. Mag.*, 1919.
Potassium ferrocyanide, $K_4Fe(CN)_6 \cdot 3H_2O$ . .	8·408	" "
Gypsum, $CaSo_4 \cdot 2H_2O$.	7·621	" "

lines have been found to be associated with the characteristic X-ray spectrum of each element. Three series of lines are known at present—the K, L, and M, of which the K has the highest frequency. It has also been claimed that a J series exists, but the evidence needs confirmation. The K series contains at least 4 lines, α, β, γ, δ (some of them doublets), of which the δ line has the highest frequency. The L series contains probably 3 groups of lines, each group similar to the K series.

The values of the wave-lengths of the principal lines are given in the tables in Ångström units. It should be noted that all the values rest on W. L. Bragg's estimate of the lattice-constant of rock salt (see above).

(iii.) *Absorption Coefficients.*—A very usual method of determining the quality of X-rays is to measure their absorption in a standard material such as aluminium. Aluminium is commonly chosen because it is readily procurable in convenient form, and, so far as is known, does not, in the majority of cases, complicate matters by superposing a characteristic radiation.

Now it is found that if all the rays, both entering and leaving a plate of material, are homogeneous (that is, wholly of the same quality), then the rays are absorbed exponentially by the plate, *i.e.* if 1, 2, 3 . . . similar sheets are successively introduced, each additional sheet absorbs the same fraction of what it receives. In other words, if there is no "scattering" or transformation of the X-rays, and if μx is the fraction of the intensity which is absorbed when the rays pass normally through a very thin screen of thickness x (cm.), then for a plate of thickness d (cm.)

$$I = I_0 \, . \, e^{-\mu d},$$

in which I_0 is the intensity of the beam when it enters, and I that of the beam when it leaves the screen. e (=2·72) is the base of the hyperbolic system of logarithms. μ is termed the linear absorption coefficient.

K Series (Principal Lines)

At. No.	Element.	α_2.	α_1.	β_1.	β_2.	Observer.
		$\times 10^{-8}$ cm.	$\times 10^{-8}$ cm.	$\times 10^{-8}$ cm.	$\times 10^{-8}$ cm.	
11	Na	..	11·95	..	..	Siegbahn and Stenström, *P.Z.*, July 1916.
12	Mg	..	9·92	9·48	..	,, ,,
13	Al	..	8·36	7·99	..	,, ,,
14	Si	..	7·13	6·76	..	,, ,,
15	P	..	6·17	5·81	..	,, ,,
16	S	..	5·36	5·02	..	,, ,,
17	Cl	..	4·7187	4·39	..	Siegbahn, *P.M.*, June 1919.
19	K	..	3·7339	3·4474	..	,, ,,
20	Ca	..	3·3519	3·0879	..	,, ,,
21	Sc	..	3·0253	2·7745	..	,, ,,
22	Ti	2·746	2·742	2·509	2·492	Siegbahn and Stenström, *P.Z.*, July 1916.
23	V	2·502	2·498	2·281	..	,, ,,
24	Cr	..	2·2852	2·0814	..	Siegbahn, *P.M.*, June 1919.
25	Mn	2·097	2·093	1·902	1·892	Siegbahn and Stenström, *P.Z.*, Feb. 1916.
26	Fe	..	1·9324	1·7540	..	Siegbahn, *P.M.*, June 1919.
27	Co	..	1·7852	1·6176	..	,, ,,
28	Ni	..	1·6547	..	..	,, ,,
29	Cu	..	1·5374	1·3895	..	,, ,,
30	Zn	1·437	1·433	1·294	1·281	Siegbahn and Stenström, *P.Z.*, Feb. 1916.
32	Ge	1·261	1·257	1·131	1·121	,, ,,
39	Y	..	0·833	..	..	Moseley (corrected), *P.M.*, April 1914.
40	Zr	..	0·790	..	..	,, ,,
41	Nb	..	0·746	..	..	,, ,,
42	Mo	..	0·717	..	..	,, ,,
44	Ru	..	0·635	..	..	,, ,,
45	Rh	0·6164	0·6121	0·5453	0·5342	Duane and Hu, *P.R.*, 1919.
46	Pd	0·589	0·583	0·516	..	Bragg.
47	Ag	0·562	0·557	0·495	..	,,
48	Cd	..	0·537	0·475	..	Siegbahn, *D.P.G.V.*, 1916.
49	In	..	0·506	0·454	..	,, ,,
50	Sn	..	0·486	0·432	..	,, ,,
51	Sb	..	0·469	0·416	..	,, ,,
52	Te	..	0·456	0·404	..	,, ,,
53	I	..	0·437	0·388	..	,, ,,
56	Ba	..	0·388	0·344	..	,, ,,
74	W	0·2135	0·2089	0·1844	0·1794	Siegbahn, *P.M.*, Nov. 1919.
92	U	..	0·15	0·10	..	,, ,,

L Series (Principal Lines)

At. No.	Element.	α_2.	α_1.	β_1.	β_2.	γ_1.	Observer.
		$\times 10^{-8}$ cm.	$\times 10^{-8}$ cm.	$\times 10^{-8}$ cm.	$\times 10^{-8}$ cm.	$\times 10^{-8}$ cm.	
30	Zn	..	12·35	..	..	..	Friman, *P.M.*, Nov. 1916.
33	As	..	9·701	9·449	..	..	,, ,,
35	Br	..	8·391	8·141	..	..	,, ,,
37	Rb	..	7·335	7·091	..	..	,, ,,
38	Sr	..	6·879	6·639	..	..	,, ,,
39	Y	..	6·464	6·227	..	..	,, ,,
40	Zr	..	6·083	5·851	..	5·386	,, ,,
41	Nb	5·731	5·724	5·493	5·317	..	,, ,,
42	Mo	5·410	5·403	5·175	..	..	,, ,,
44	Ru	4·853	4·845	4·630	..	..	,, ,,
45	Rh	..	4·596	4·372	..	..	,, ,,
46	Pd	4·374	4·363	4·142	3·903	3·720	,, ,,
47	Ag	4·156	4·146	3·929	3·698	3·515	,, ,,
48	Cd	3·959	3·949	3·733	3·514	3·331	,, ,,
49	In	3·774	3·766	3·550	3·335	3·160	,, ,,
50	Sn	3·604	3·594	3·381	3·172	2·999	,, ,,

D.P.G.V., *Verh. der Deutsch. Phys. Gesell.*; *P.M.*, *Phil. Mag.*; *P.R.*, *Phys. Rev.*

L SERIES (PRINCIPAL LINES)—*continued.*

At. No.	Element.	α_2.	α_1.	β_1.	β_2.	γ_1.	Observer.
		$\times 10^{-8}$ cm.	$\times 10^{-8}$ cm.	$\times 10^{-8}$ cm.	$\times 10^{-8}$ cm.	$\times 10^{-8}$ cm.	
51	Sb	3·443	3·434	3·222	3·021	2·849	Friman, *P.M.*, Nov. 1916.
52	Te	3·299	3·290	3·074	2·881	2·712	,, ,,
53	I	3·155	3·146	2·934	2·750	2·583	,, ,,
55	Cs	2·899	2·891	2·684	2·514	2·350	,, ,,
56	Ba	2·786	2·776	2·569	2·407	2·245	,, ,,
57	La	2·674	2·665	2·461	2·307	2·146	,, ,,
58	Ce	2·573	2·563	2·359	2·212	2·052	,, ,,
59	Pr	2·472	2·462	2·259	2·120	1·958	,, ,,
60	Nd	2·379	2·369	2·167	2·036	1·875	,, ,,
62	Sa	2·210	2·200	2·000	1·884	1·725	,, ,,
63	Eu	2·131	2·121	1·918	1·810	1·662	,, ,,
64	Gd	2·054	2·043	1·844	1·744	1·597	,, ,,
65	Tb	1·983	1·973	1·775	1·682	1·531	,, ,,
66	Dy	1·916	1·907	1·709	1·622	1·470	,, ,,
67	Ho	1·854	1·843	1·646	1·568	1·415	,, ,,
68	Er	1·794	1·783	1·586	1·514	1·367	,, ,,
70	Yb	1·681	1·670	1·474	1·414	1·267	,, ,,
71	Lu	1·629	1·619	1·421	1·368	1·224	,, ,,
73	Ta	1·528	1·518	1·323	1·280	1·135	Siegbahn and Friman, *P.M.*, July 1916.
74	W	1·4845	1·4735	1·2792	1·2419	1·0955	Siegbahn, *P.M.*, Nov. 1919.
76	Os	1·398	1·388	1·194	1·167	1·021	Siegbahn and Friman, *P.M.*, July 1916.
77	Ir	1·360	1·350	1·154	1·133	0·980	,, ,,
78	Pt	1·323	1·313	1·120	1·101	0·958	,, ,,
79	Au	1·283	1·271	1·080	1·065	0·922	,, ,,
80	Hg	1·251	1·240	1·049	1·042	0·896	,, ,,
81	Tl	1·215	1·205	1·012	1·006	0·864	,, ,,
82	Pb	1·186	1·175	0·983	0·983	0·842	,, ,,
83	Bi	1·153	1·144	0·950	0·954	0·810	,, ,,
84	Po	..	1·109	0·920	..	..	,, ,,
88	Ra	..	1·010	..	..	..	,, ,,
90	Th	0·969	0·957	0·766	0·797	0·654	,, ,,
92	U	0·922	0·911	0·720	0·756	0·615	,, ,,

M SERIES (PRINCIPAL LINES)

At. No.	Element.	α.	β.	γ.	δ.	Observer.
		$\times 10^{-8}$ cm.	$\times 10^{-8}$ cm.	$\times 10^{-8}$ cm.	$\times 10^{-8}$ cm.	
79	Au	5·838	5·623	5·348, 5·284	5·146, 5·102	Siegbahn, *D.P.G.V.*, 1916.
81	Tl	5·479	5·256	..	4·826	,, ,,
82	Pb	5·303	5·095	4·91(?)	4·695	,, ,,
83	Bi	5·117	4·903	4·726, 4·561	4·532, 4·456	,, ,,
90	Th	4·139	3·941	3·812, 3·678	..	,, ,,
92	U	3·905	3·715	3·480	3·363, 3·324	,, ,,

D.P.G.V., Verh. der Deutsch. Phys. Gesell.; *P.M., Phil. Mag.*.

It follows that $\mu=(2{\cdot}3/d)(\log I_0-\log I)$; the logarithms being to base 10. If in a set of observations with homogeneous rays, log I is plotted as ordinate against d, the graph is a straight line and μ is 2·3 times the slope of the line.

With ordinary heterogeneous rays, μ is greater for thin screens than for thick, and so we can only deal with an average μ, which, however, varies more and more slowly as the screen becomes thicker (*Fig.* 7).

The logarithmic curve of absorption for

heterogeneous rays, such as are given out by an ordinary X-ray bulb, is not a straight line, but a curve which is steeper for thin screens than for thick (*Fig.* 8). For a method of finding analytically the absorption coefficients of the constituents of a complex beam of rays, see J. J. Thomson, *Phil. Mag.*, December 1915.

In the case of the characteristic radiations, an element exhibits a maximum transparency

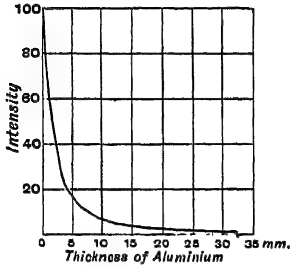


FIG. 7.

for each of its own characteristic radiations. For slightly harder rays than these, the absorption rapidly increases; the rays characteristic of the screen are produced and superposed on the transmitted rays to an extent which diminishes as the incident rays are increasingly

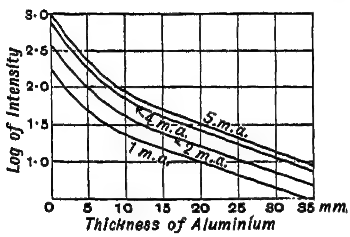


FIG. 8.

hardened. For incident rays softer than the critical type, no characteristic rays are produced. Thus, as the incident rays are gradually hardened, the transmitted rays reach a maximum intensity when the incident rays are of the same quality as each of the characteristic rays in turn.

A large value of μ corresponds to easily absorbed rays, and a small one to very penetrating rays. μ also varies with the nature of the absorbing screen, so that it is necessary to specify the material used. For medical purposes, it has recently been suggested that water should be chosen as the standard absorbing medium, since the absorptive power of water agrees closely with that of animal tissue.

Some workers prefer to think in terms of the thickness D, which reduces the intensity to half value. D is connected with μ by the expression $D=0{\cdot}69/\mu$. A notion of the order of values of μ may be got from the fact that for an X-ray beam of average hardness μ_{Al} lies between 4 and 8 cm.$^{-1}$; for hard rays between 2 and 4 cm.$^{-1}$. μ for fatty tissue varies from about 0·4 for hard rays to 0·7 for medium rays.

A more fundamentally important constant is obtained by dividing the absorption coefficient (μ) by the density (ρ) of the absorbing screen. This quantity, μ/ρ—usually called the mass-absorption coefficient—gives a measure of the absorption per unit mass of the screen for a normally incident pencil of rays of unit cross-section. Since it is mass alone that affects absorption, at any rate as determined by the usual methods of measurement, it is ordinarily more convenient to use mass coefficients than linear coefficients.

If, as was at one time supposed, the absorbing powers of different materials were truly proportional to their densities, then for the same rays μ/ρ would be a constant, no matter what the substance used as screen. In point of fact, dense substances are a good deal more absorptive, mass for mass, than light, and μ/ρ increases rapidly with the atomic weight of the screen. The increase is more noticeable with hard rays than with soft.

The accompanying table gives a series of values connecting wave-lengths and absorption coefficients in aluminium, derived from the results of Rutherford, Bragg, Moseley, and Barkla. A scrutiny of these results shows that if μ is the absorption coefficient and λ is the wave-length, then, when the effect of scattering has been allowed for,

$$\mu=k\lambda^n,$$

where k is a constant, and n lies between 5/2 and 3.

Wave-length λ.	μ/ρ_{Al}.	Wave-length λ.	μ/ρ_{Al}.
1×10^{-9} cm.	0·04	12×10^{-9} cm.	22
2 ,,	0·21	13 ,,	28
3 ,,	0·57	14 ,,	35
4 ,,	1·20	15 ,,	43
5 ,,	2·10	16 ,,	51
6 ,,	3·3	17 ,,	61
7 ,,	4·8	18 ,,	72
8 ,,	6·6	19 ,,	83
9 ,,	8·9	20 ,,	95
10 ,,	12·2	21 ,,	108
11 ,,	16·5	22 ,,	120

§ (18) THE SPEED OF THE CATHODE RAYS.—If the exciting voltage is uniform and a magnetic field be applied to the cathode rays, the band of rays is deflected as a whole to an extent dependent on the magnetic field and

the speed of the cathode rays. If the exciting voltage is pulsating, and a similar experiment be tried, a magnetic spectrum of cathode rays is formed, the least deflected band of rays corresponding to the highest speed rays which owe their velocity (v) to the maximum voltage (V) applied.

Since the energy of the electron is equal to the work done in expelling it, we have

$$\tfrac{1}{2}mv^2 = \mathrm{V}e,$$

where e and m are respectively the electronic charge and mass. Substituting the accepted value of e/m

$$\mathrm{V} = 2{\cdot}8v^2 . 10^{-16},$$

where V is in volts and v is in cm. per sec.

It is thus possible by measuring v by means of the magnetic deflection in a known field to arrive at the value of V.

§ (19) THE "QUANTUM LIMIT" TO THE X-RAYS.—The X-ray spectrometer has brought out the truth of a remarkably simple relation between the voltage applied to an X-ray bulb and the shortest wave-length emitted. It is now well known that, no matter whether the exciting voltage is constant or pulsating, a spectrum of X-rays is generated containing a wide range of wave-lengths. This spectrum is terminated very definitely at the short-wave

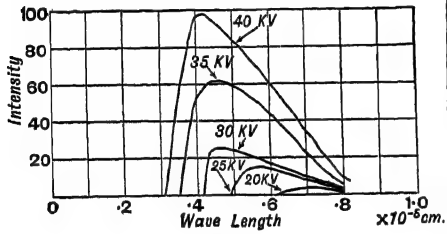


FIG. 9.

end (*Fig.* 9) at a point determined by Planck's quantum[1] relation

$$\mathrm{V}e = h\nu,$$

where

V is the maximum voltage applied,
e is the electronic charge,
h is Planck's universal constant,
ν is the frequency of the limiting wave.

Substituting the accepted values of the constants in the above equation, we have

Voltage = 12,400/wave-length in A.U.

It is obvious that the X-ray spectrometer can thus be utilised in some convenient form as a means of measuring the maximum effective voltage applied to a tube.

§ (20) THE MAXIMUM SPARK GAP.—The maximum voltage applied to a tube is most commonly measured by means of the alternative spark gap between points or spheres. Some experience is necessary, especially in the case of a gas X-ray tube, where the method tends to give too high values, especially with pulsating potentials. In the case of Coolidge radiator tubes excited by unrectified alternating potential, the spark gap will register the "inverse" voltage rather than the lower "effective" voltage.

The work of Peek and others has shown that a spark gap between spherical electrodes of equal size is preferable to that between points. The spark between points is now generally discredited for high voltages on account of its inconsistent dependence on atmospheric humidity and frequency of discharge. By reason of its time-lag, its readings may be largely in error, in the case of high-frequency steep impulses.

On the other hand, frequency and wave shape have no appreciable effect in the case of the sphere gap, and the effects of variation in the atmospheric conditions are well known, and can be readily corrected for.

The size of the spheres is important. A good rule is not to use a gap bigger than the diameter of either of the balls, though some latitude may be permitted in this direction. The main point is to avoid the break-down discharge being preceded by brush-discharge or corona, otherwise a pulsating discharge will, in general, give gap readings much too high. With the above precaution, a sphere gap is capable of measuring (peak) voltages from, say, 10,000 volts to 500,000 to an accuracy of about 2 per cent.

The table below is based on Dr. A. Russell's formula and incorporates the latest results of the American Institute of Electrical Engineers (1918). It includes also, for convenience, a column of figures for a needle point gap (No. 00, new sewing needles) which furnish a rough notion of the voltages for an instrument which is still much used. The A.I.E.E. recommend that for voltages above 70,000 (and preferably above 40,000) a sphere gap should always be employed.

The gap should not be exposed to any extraneous ionising influence, such as an arc or an adjacent spark, nor should the gap be enclosed. The first spark is the one for which the reading should be taken. The use of a water resistance in series with the gap will prevent arcing and pitting of the sphere surfaces.

§ (21) SPARK-GAP VOLTAGES AT 760 MM. PRESSURE AND 25° C.—Where any gap is being used outside its recommended limits, the figures are shown in brackets. The blank spaces indicate that the gap is no longer suitable. The gaps are given to three significant figures for interpolation purposes.

[1] See "Quantum Theory," § (1).

TABLE A

Kilo Volts (peak).	Diameter of Spheres.						
	Needle Points.		2·5 cm.	5 cm.	10 cm.	25 cm.	50 cm.
	cm. gap.	inches gap.	cm. gap.	cm. gap.	cm. gap.	cm. gap.	cm. gap.
5	(0·42)	(0·17)	(0·13)	(0·15)	(0·15)	(0·16)	(0·17)
10	(0·85)	(0·33)	0·27	0·29	0·30	0·32	0·33
15	1·30	0·51	0·42	0·44	0·46	0·48	0·50
20	1·75	0·69	0·58	0·60	0·62	0·64	0·67
25	2·20	0·87	0·76	0·77	0·78	0·81	0·84
30	2·69	1·06	0·95	0·94	0·95	0·98	1·01
35	3·20	1·26	1·17	1·12	1·12	1·15	1·18
40	3·81	1·50	1·41	1·30	1·29	1·32	1·35
45	4·49	1·77	1·68	1·50	1·47	1·49	1·52
50	5·20	2·05	2·00	1·71	1·65	1·66	1·69
60	6·81	2·68	2·82	2·17	2·02	2·01	2·04
70	8·81	3·47	(4·05)	2·68	2·42	2·37	2·39
80	(11·1)	(4·36)	..	3·26	2·84	2·74	2·75
90	(13·3)	(5·23)	..	3·94	3·28	3·11	3·10
100	(15·5)	(6·10)	..	4·77	3·75	3·49	3·46
110	(17·7)	(6·96)	..	5·79	4·25	3·88	3·83
120	(19·8)	(7·81)	..	(7·07)	4·78	4·28	4·20
130	(22·0)	(8·65)	..	..	5·35	4·69	4·57
140	(24·1)	(9·48)	..	..	5·97	5·10	4·94
150	(26·1)	(10·3)	..	..	6·64	5·52	5·32
160	(28·1)	(11·1)	..	..	7·37	5·95	5·70
170	(30·1)	(11·9)	..	..	8·16	6·39	6·09
180	(32·0)	(12·6)	..	..	9·03	6·84	6·48
190	(33·9)	(13·3)	..	..	10·0	7·30	6·88
200	(35·7)	(14·0)	..	..	11·1	7·76	7·28
210	(37·6)	(14·8)	..	..	(12·3)	8·24	7·68
220	(39·5)	(15·5)	..	..	(13·7)	8·73	8·09
230	(41·4)	(16·3)	..	..	(15·3)	9·24	8·50
240	(43·3)	(17·0)	..	..	..	9·76	8·92
250	(45·2)	(17·8)	..	..	..	10·3	9·34

§ (22) CORRECTION FOR DENSITY OF AIR.—*Applicable only to sphere gaps.* The following table gives the relative air density under different conditions. The figures are relative to dry air at 25° C. and 760 mm. pressure :

TABLE B

Temperature ° C.	Pressure 720 mm.	Pressure 740 mm.	Pressure 760 mm.	Pressure 780 mm.
0	1·04	1·06	1·09	1·12
10	1·00	1·02	1·05	1·08
20	0·96	0·99	1·02	1·04
30	0·93	0·96	0·98	1·01

Within the limits of the above table the correction factor for a sphere gap agrees substantially with the relative air density. Thus for a given length of spark gap, the tabulated kilo-voltage in Table A must be multiplied by the appropriate correction factor in Table B. Alternatively, to calculate the gap which will just be sparked over by some specified voltage, the voltage must first be divided by the appropriate correction factor before Table A is used.

It will be seen that under normal conditions the correction is small or negligible.

§ (23) PENETROMETERS. (i.) *Benoist Penetrometer.*—Among medical men Benoist's radiochromometer or penetrometer enjoys extensive use as a measurer of hardness. It consists of a thin silver disc 0·11 mm. thick, surrounded by twelve numbered aluminium sectors from 1 to 12 mm. thick. The X-rays are sent through the instrument, and the observations consist merely in matching on a fluorescent screen or photographic plate the image cast

by the silver disc against the images of the aluminium plates: the thickness of the matching sector increases with the hardness of the rays. A notion of the discharge potential across a tube may be got from the very rough relation that the voltage is from 8000 to 10,000 times the Benoist reading of the X-rays.

(ii.) *Walter Penetrometer.*—This consists of a number of holes in a lead disc, which are covered by a sequence of platinum discs of gradually increasing thickness.

(iii.) *Christen's Half-value Penetrometer.*—Christen adopts as a definition of quality the thickness of a layer of water (or, in actual practice, bakelite), which will reduce the intensity of a beam of rays to half its original value.

The rays are sent through a stepped wedge of bakelite, alongside which is a perforated metal plate. This provides a standard of reference on a fluorescent screen, the two images being side by side. The holes in the plate are so designed that the area of the metal removed equals that which remains, so that the plate by this means reduces the intensity of a beam of rays to half-value. The holes are small enough to produce uniform illumination on a screen placed a short distance behind the plate.

§ (24) Qualimeters.—(i.) *Bauer Qualimeter.*—This is a type of semi-electrostatic voltmeter, which serves to measure the potential difference between the electrodes of a tube.

(ii.) *Klingelfuss Qualimeter* consists of an auxiliary search coil and electrostatic voltmeter. The instrument works similarly to the Bauer.

The hardness-numbers of the various penetrometers are all much the same as Benoist's, except those of Wehnelt, which are 50 per cent bigger for the same quality of rays.

§ (25) Methods of Measuring Intensity.—The intensity of the X-rays at a particular point is defined as the energy falling on one square centimetre of a receiving surface passing through the point and placed at right angles to the surface. The question is thus on all fours with that of illumination with visible light, and the need of a unit of "candle-power" in X-ray work is becoming pressing.

The work of a number of experimenters with the X-ray spectrometer has shown that the energy E emitted per second by an X-ray bulb may be written

$$E = kNiV^2,$$

where

V is the voltage on the tube,
i is the current through the tube,
N is the atomic number of the target,
k is a constant.

The value of k will depend on whether a gas or a hot-cathode tube is employed, and also on the type of exciting potential. The above expression refers only to the general or "white" X-radiation. If the voltage is such as to excite the characteristic radiations, the voltage comes in as a higher-powered term than V^2, and the efficiency increases correspondingly.

Measurements of the value of k have been made, and result in showing, unfortunately, that the efficiency of an X-ray bulb is deplorably small, of the order of 1 part in 1000. The chances that a cathode ray will ultimately come into suitable conflict with some atom and so generate an X-ray are slight. We raise those chances by increasing either the exciting voltage or the mass of the atom of the target.

The X-ray emission is virtually distributed over a sphere, or, more practically, over a hemisphere, since the target blocks out most of one half of the sphere of radiation.

Thus the intensity (comprising all wavelengths) falling on a square centimetre at distance d from the anticathode may be written

$$\frac{kNiV^2}{d^2}.$$

If the length of exposure is t secs. the total amount of energy received becomes

$$\frac{kNiV^2t}{d^2}.$$

It is apparent that we can base on this relation a system of X-ray intensities and "doses," provided we can measure i and V and be certain to what extent each is effective in generating rays which are of practical utility.

In practice we almost always filter out the long waves, and we shall need then to know the new value of k, so as to correct for the proportion of i which is not usefully employed.

The measurement of all the terms in the above expression, except perhaps i and V, offers no difficulty. i can be measured by a milliammeter of suitable design, which even with rapidly pulsating currents appears to indicate the mean current, as has been verified by the use of the voltameter and oscillograph.

The measurement of V has already been referred to above. If the exciting potential is constant, it can be measured by a resistance type of voltmeter in series with a high resistance. If the potential is fluctuating, recourse will usually be had to the sphere spark gap. The voltages are mostly too high for an electrostatic voltmeter.

In addition to the above method, the radiologist has employed a variety of means of measuring the intensity of X-rays at some selected point in the beam. To this end one or other of the properties of the rays have been utilised—heating, ionising, photographic, fluorescing, or chemical.

The heating effects are minute and the method is only fitted for the research laboratory.

(i.) *Ionisation Methods.*—When X-rays pass through a gas they impart to it a temporary electrical conductivity, the extent of which depends on the number of ions formed and thus on the amount of energy absorbed in the gas. An ionisation method of evaluating X-rays thus resolves itself into the measurement of a minute electric current. For this purpose an electroscope or electrometer is commonly employed, and the delicacy and convenience are such that the method has found almost universal acceptance among research workers.

But in medical radiology the method is only beginning to find favour, more especially in deep therapy. The iontoquantimeter of Szilard was one of the first instruments to be designed on this basis. In its most recent development a part of the quantimeter is sometimes actually introduced into the affected organ, and the rays measured which are actually received at the point concerned.

The ionisation method is unapproached in sensitiveness by any other method; it does not depend dominantly upon any selective process, and it is reasonable to anticipate that some unit of dosage thus developed and connected maybe with the accepted radium standard will ultimately receive recognition as a standard.

(ii.) *Photographic and Fluorescence Methods.*—Photographic plates record only about 1 per cent of the energy of the X-rays, but nevertheless a method of measuring intensity by this means has been developed and finds favour with some workers.

As is the case with most types of intensity meters the short-waved rays are recorded disadvantageously compared with the soft rays. Furthermore, a photographic film betrays marked selective absorption, and the photographic effect may be quite misleading in consequence.

To the worker with limited resources the photographic method of measuring intensity offers advantages because of its simplicity. Some form of opacity-meter for obtaining a measure of the density of the image is the chief requirement. The opacity meter measures the extent to which a standard beam of light is cut down by the photographic film whose density is required. If I_0 is the intensity of the (homogeneous) testing light which is incident on the developed film, and I_t that of the transmitted light, then, if μx is the fraction of the energy which is absorbed by a very small thickness, x, of the film,

$$I_t = I_0 e^{-\mu d},$$

where d is the thickness of the film. The film is assumed to be equally dense throughout its thickness.

For films of uniform thickness, d is constant, so that μ is proportional to $\log (I_0/I_t)$. μ is called the absorption coefficient; (I_0/I_t) is known as the opacity, and measures the number of times the incident light is cut down. $\operatorname{Log}(I_0/I_t)$ is termed the opacity logarithm. The transparency is the reciprocal of the opacity. Now, by definition, μ is proportional to the density of the image—*i.e.* to the amount of silver per unit area of film. Thus the ratio of two opacity logarithms gives the ratio of the film densities, and therefore the ratio of the photographic energies in the two cases. The opacity meter is graduated to read directly in opacity logarithms.

In fluorescence methods the luminosity is matched against some standard fluorescence excited by a steady source of radiation such as radium. The drawback to such methods is that the fluorescing salt becomes "tired" under the action of the rays. The sensitivity of a screen may also vary considerably from point to point, so that it is difficult to make a fair comparison. Barium platinocyanide is the material commonly used to sensitise a fluorescent screen. This salt, which has the formula $BaPt(CN)_4$, $4H_2O$, exists in three different forms, of which the green crystalline variety is by far the most efficient for fluorescing purposes.

(iii.) *Chemical and Other Methods.*—In the therapeutic use of X-rays, various chemical reactions brought about by the rays have been suggested and employed from time to time as aids to "dosage"; for example, the discolouring of various alkaline salts (Holzknecht, 1902); the liberation of iodine from a 2 per cent solution of iodoform in chloroform (Freund, 1906; Bordier and Galimard, 1906); the darkening of a photographic paper (Kienböck); the precipitation of calomel from a mixture of mercuric chloride and ammonium oxalate solutions (Schwarz, 1907); and the change of colour of pastilles of compressed barium platinocyanide (Sabouraud-Noiré and Bordier). X-rays resemble light in their property of lowering the electrical resistance of selenium; this property, if the pronounced fatiguing of the selenium could be overcome, would doubtless furnish the basis of a very convenient method of measurement, though Fürstenau in his intensimeter claims to have got over this difficulty. It must be admitted that most of these methods provide little more than the roughest notion of the intensity of a beam of ordinary heterogeneous X-rays.

Of all the various intensity-measurers, the pastille finds most favour with medical men. The barium-platinocyanide discs are some 5 mm. in diameter, and their colour, initially a bright green, changes, when exposed to the rays, to a pale yellow, and finally to a deep orange. The pastille is placed at a specified

distance from the anticathode of the bulb, and the colour is matched against one of a number of standard tints. The method is easy in practice, and is fairly reliable as a guide for short exposures, but it is not trustworthy for heavily filtered or very penetrating rays. The pastille method is defective in that it attempts to measure rays of all qualities by a surface coloration. Other platinocyanides show similar colour changes when exposed to X-rays. Levy has shown that the change of colour is due to a change from the crystalline to the amorphous condition. If the pastille is put aside, the reverse change slowly takes place, especially in the presence of light, so that the pastille should not be exposed to full daylight during the X-ray treatment. Ultra-violet light and radium rays cause similar browning in such pastilles.

The following table gives an idea of the relation between the different dosemeter scales:

5H units (Holzknecht; alkaline salt)
=Tint B (Sabouraud-Noiré; pastille)
=Tint 1 (Bordier; varnished pastille)
=3 to 4I (Bordier and Galimard; iodine solution)
=10X units (Kienböck; photographic plate)
=3·5 kaloms (Schwarz; mercury solution)
=Villard dose.

§ (26) Effect of Wave Form of Exciting Potential.—So long as radiologists confined their measurements on the X-ray bulb to the milliammeter and the point gap, little progress was made on the subject of the best form of exciting potential wave. It was realised that the milliammeter was often misleading if used alone, but insight into the problem only came with the use of the X-ray spectrometer and the oscillograph.

The oscillograph (preferably of the cathode ray type) should be arranged to give simultaneous graphs of both potential and current wave forms. The spectrometer analyses the resulting X-ray output and gives the distribution of intensity among the various wave-lengths.

Work such as this has shown the truth of what might have been anticipated on *a priori* grounds, *i.e.* the importance in pulsating discharges of keeping the conditions so that the potential and current curves are in phase. In other words, for efficient output of X-rays as much of the current as possible should be sent through the tube with the maximum potential actuating it.

As the output of X-rays depends on the square of the voltage, the ideal state of things would appear to be to rush the potential as quickly as possible to the maximum and keep it there for as long as any current is passing. Voltages lower than the maximum are less efficient.

In practice much depends on the type of X-ray bulb used—whether gas or hot cathode, and it will be instructive to compare the behaviour of three main types of exciting potentials (*a*) constant potential, (*b*) sinusoidal transformer discharge, (*c*) sharp-peaked pulsating coil discharge, all running under the same maximum voltage and the same milliamperage.

First take the case of the Coolidge tube. Owing to the existence of the saturation current, the shape and limits of the current curve are greatly dependent on the characteristics of the potential wave (*Fig.* 10).

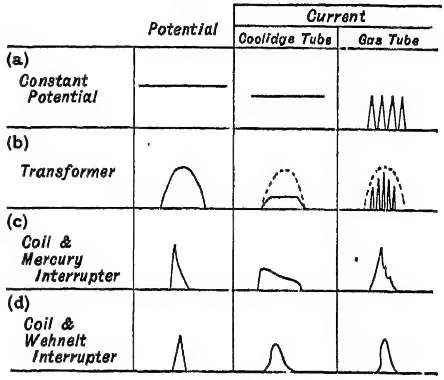


Fig. 10.

With constant potential, the current remains constant at its saturation value. In the case of the transformer loop the current rises gradually to its saturation value and dies away again as the potential loop terminates. In the case of the coil and interrupter, the current again rises to its saturation value, but as a whole the current curve is prone to lag behind the potential curve. The defect, which is due to arcing within the interrupter, increases as the current through the tube is increased. This is the explanation of the well-known fact that as the current is raised the output of X-rays increases only slowly and by no means proportionately. A partial cure is to raise the speed of the interrupter, as in the Wehnelt break.

We are led to anticipate the results of the spectrometer curves. Remembering that the area of each curve is a measure of the total output of radiation, we see from *Fig.* 11, which gives intensity in terms of wave-length (derived from Dauvillier), that the constant potential is the most efficient, then the transformer, and lastly the coil. Furthermore, the crest of the curve is of slightly shorter wave-length in the case of the constant potential. It may be added that the superiority of the constant potential becomes less marked as the potential rises.

If we now consider the case of the gas tube, the current curves prove, in most circum-

stances, to be quite different from those for the Coolidge tube (*Fig.* 10). The current curve now exhibits no saturation value, but consists instead of a succession of peaks, no

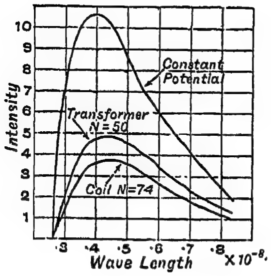


FIG. 11.

matter whether the exciting potential is constant or varying. The constant potential produces, so to speak, all the effects of an intermittent one. In the case of the transformer the middle portion of the potential loop produces a number of current peaks whose heights wax and wane with the potential. The coil produces one or more current peaks corresponding to each potential peak. We are led to infer that the spectral curves for the different excitants will not differ appreciably, and this proves to be the case. *Fig.* 12 gives the curves derived from three

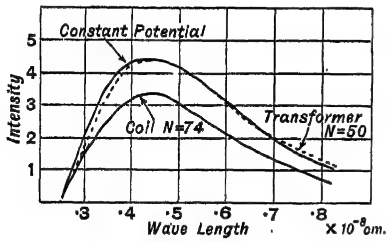


FIG. 12.

different excitants. Further, the maxima of the several spectral curves all agree in wavelength. We realise then that the shape and limits of the spectral curve are determined mainly by the tube rather than by the excitant.

In the case of both gas and hot cathode tubes there is a variety of factors which may modify the various curves and graphs appreciably in detail, without affecting the main outlines.

To sum up, we may conclude that for maximum efficiency in the case of a Coolidge tube we should raise the potential on the tube as high as is expedient or possible and keep it there. Thus the ideal excitant would appear to be constant potential or, failing that, a transformer of high frequency.

The "shock tactics" of a coil, which are largely wasted on a Coolidge tube, are doubtless well adapted to a gas tube owing to its characteristic way of breaking down under potential stress. From a practical point of view, the transformer potential wave is probably not quite so well suited to a gas tube, but constant potential should prove very effective apart from difficulties of generation, etc. The practical difficulties of obtaining an equally high constant potential or maximum transformer voltage (selectively rectified) are probably greater, but otherwise they would seem to offer advantages equal to those of the coil.

The interval between the discharges of a coil-driven bulb is not without its advantage in helping to keep down the temperature rise of the target. But with a mercury break under ordinary conditions this interval amounts to about 90 per cent of the time between successive impulses, and is needlessly long. In the case of a single-phase transformer the corresponding figure is of the order of 60 per cent.

We can raise the efficiency of the coil or transformer by increasing the frequency substantially, and so crowd in more impulses per second, though the increased heating of the target may have to be met. If, further, in the case of a coil, we raise the voltage on the primary and increase the capacity of the condenser, we can produce a series of flat-topped peaks with little or none of the "tail" in evidence. We are thus approximating more and more to the transformer (selectively rectified) and, in the limit, the ideal steady potential. During these changes the readings of the milliammeter will approach more and more the effective values from an X-ray point of view.

§ (27) CONSTANT POTENTIAL.—The value of constant potential has been referred to. It is not only an efficient means of generating X-rays, but it permits precise measurement. As already remarked, the superiority of constant over varying potential is less marked as the voltage is increased, except when the characteristic radiations begin to be generated, when the constant potential increases its relative effectiveness. Moreover, constant potential is admirable in its precision for therapeutic purposes. In radiography it generates sufficient diversity of wave-lengths to give good contrast and detail.

Those workers who have used the influence machines in America and elsewhere speak highly of the results. There is a great future for a static machine of engineering design and large output which will defy the varying and generous humidity of this country.

The only other means of obtaining steady potential are by use of the transformer

together with the Kenetron or hot-cathode valve. These latter can now be obtained to rectify 100,000 volts. By the use of 3-phase current and 6 kenetrons a voltage fluctuating only 15 per cent can be obtained. There are a variety of ways of combining kenetrons with

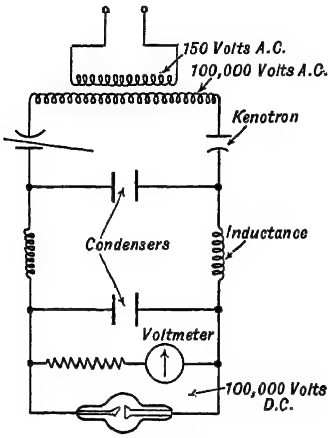


FIG. 13.

condensers and inductances, so that the variation in the resulting potential is trifling. For example (*Fig.* 13), Hull of the G.E.C. Laboratory at Schenectady has so transformed and rectified 150 volts A.C. ($\omega = 2000$) to 92,000

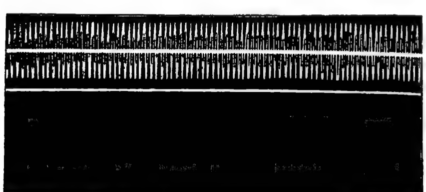


FIG. 14.—Upper Curve unrectified, Lower Curve rectified.

with a fluctuation of about 1 per cent (*Fig.* 14); or 50,000 volts with a fluctuation of only $\frac{1}{10}$ per cent.

§ (28) HOMOGENEOUS X-RAYS.—An ideal of the radiologist has always been a means of producing homogeneous X-rays, so that, among other things, a precision dose can be the better formulated in therapy. While all X-ray beams are heterogeneous, they tend to become less so as the voltage is raised and the rays are filtered by the right choice of substance of a suitable thickness. This may be seen either from absorption curves or, more accurately, from spectral distribution curves.

But the nearest approach to homogeneity can be reached by operating the X-ray bulb at a voltage somewhat above one of the critical values necessary to generate one of the three known characteristic radiations (K, L, M) of the anticathode, and then filtering by a screen either of the same element as the anticathode or, preferably, by an element of slightly smaller atomic number. We have already remarked that above the critical voltage the characteristic radiations are generated more copiously than the general radiation. The voltage should not be too high, however, or short-wave general radiation will begin to make its appearance and will not be removed by filtration. There is, in fact, an optimum voltage. For example, while the critical voltage for the K radiation of tungsten (At. No. = 74) is about 70,000 volts, the optimum voltage is about 100,000 volts A filter of tungsten or tantalum (At. No. = 73), 0·15 mm. thick (*Fig.* 15; Hull),

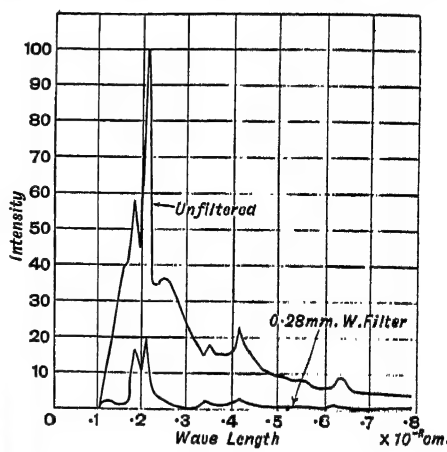


FIG. 15.

removes most of the general radiation, but leaves both the β and α components of the K radiation. An equally thick filter of ytterbium (At. No. = 70) would leave only the α component, with an intensity at least 30 times that of the remaining general radiation. Similarly for molybdenum (At. No. = 42) the optimum voltage is about 30,000 and the best filter zirconium (At. No. = 40) (see *Fig.* 16; Hull). Tungsten would be a good filter for platinum radiation (At. No. = 78).

At the National Physical Laboratory there is a battery of X-ray tubes each with a different anticathode, so that a variety of homogeneous rays can be obtained. Unfortunately such beams are of feeble intensity.

V. X-RAYS AND MEDICINE

§ (29) RADIOGRAPHY.—X-rays have become one of the handmaidens of medicine, a fact which is bound up with the great improve-

ments of recent years in X-ray equipment and technique, which have given the diagnostic methods of physicians and surgeons a facility and exactness never dreamt of at one time. Exposures have been enormously shortened (*Figs.* 17 and 18), and snapshots of any part of the body can now be taken.

In surgery of the bone, not only fractures, but the intimate lamellar structure of the bone can be examined; we have, moreover, learned that tumours and cysts in bones are not specially rare, and that nearly all sprains are accompanied by slight fractures of the bone. When a bone is badly splintered the dead bone splinters can be sorted out from the living.

Tumours in any part of the head can be detected and their position determined. The diagnosis and location of diseases of all parts of the alimentary canal are routine—stricture of the oesophagus, stomachic disorders, diseases of the appendix and colon, etc., can all be demonstrated by the X-rays with the assistance of special food containing matter opaque to the rays.

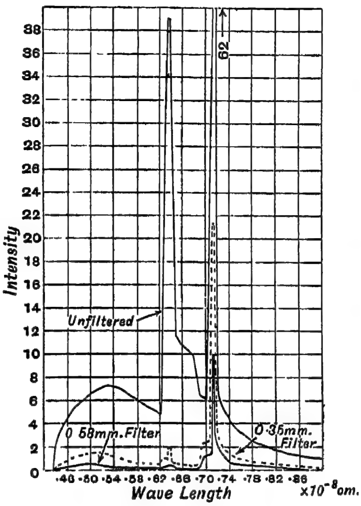


FIG. 16.

FIG. 17.—Radiograph of Hand (January 1896); Exposure 20 min. (Campbell-Swinton).

Stones in the kidney and (more recently) the gall-bladder, diseases of the liver and pelvic organs, incipient tuberculosis in the lungs and joints can be diagnosed with certainty and without pain or danger.

Dental radiography has become an important subject. By suitably disposing a photographic film radiographs of individual teeth can be obtained, revealing in perfect fashion the condition of both the tooth and surrounding bone.

During the war many thousands of radiologists helped to build up the triumphs which X-rays achieved. The X-ray became as indispensable as the dressing or the splint, and it was an essential adjunct in prescribing and directing as well as avoiding operations. The detection of bullets and shell fragments in any part of the body was commonplace, and the X-rays were also used to guide the surgeon during his actual efforts to remove foreign bodies. Cleverly designed X-ray motor-lorries permitted early examination in the field. In the case of eye wounds X-ray

FIG. 18.—Radiograph of Hand (January 1920); Exposure $\frac{1}{100}$ sec. (Knox). Showing immense advance in technique, although exposure only $\frac{1}{120000}$ of exposure in *Fig.* 17.

stereoscopy attains its fullest delicacy, and the location of small foreign bodies can be carried out to the hundredth part of an inch. Another war development of radiology was its employment by the orthopaedic surgeon in his efforts to restore damaged limbs. Many hundreds of thousands of radiographs were taken at the various hospitals during the war.

§ (30) Radiotherapy.—The X-rays possess valuable properties in the treatment of malignant disease. The living cells have the power of resisting or responding to X-rays, while malignant cells disappear with suitable "dosage." The treatment has, for example, been largely and successfully employed for rodent ulcers, and much attention is being paid to the cure of cancer.

In many skin diseases the X-rays have proved to be of notable service. For example, they are now the accepted and certain means of curing ringworm. The "dose" is all-important, for the sweat glands and hair follicles are also affected and, with excessive exposure, may even be destroyed, the result being baldness.

The corpuscles of the blood are prejudicially affected by X-rays, resulting in a form of aplastic anaemia. Curiously enough, the rays seem to have little or no action on bacteria or their spores, and in this respect stand out in marked contrast to ultra-violet light.

Deep-seated organs are now frequently treated by X-rays. In deep radiotherapy modern technique is tending in the direction of administering massive doses of X-rays in the shortest possible time, always having regard to the safety of the overlying skin. In other words, "shock tactics" are used, and the wave form of the exciting potential may prove to be important in this connection. The soft rays are removed by using metal filters (Al or Zn) with a coating on the side nearest the skin of wood or leather, so that no characteristic radiation may play on the skin. Over-dosage of the skin is also avoided by employing multiple parts of entry, each of the various beams being properly directed at the deep affected tissue.

The chief hindrance to precise radiotherapy at the present time is probably the lack of a means of measuring the dose of radiation absorbed by the particular region concerned, especially if it be at a depth in the body. On physical grounds, at any rate, it would seem that it is only those rays which are absorbed which can produce physiological changes, and only such rays should be included when speaking of a dose. It may be, of course, that selective action is present, and that only a restricted range of wave-lengths is appropriate for the conversion of energy in the correct spot. With this reservation it would seem that the degree of reaction should be a function of the absorbed and converted energy. The problem is complicated by the lack of homogeneity of the primary beam.

Among the tragedies of the war, few were more pathetic than the ghastly disfigurements caused by shell wounds of the face and head. Fortunately it was often possible, by the wonderful grafting operations of the surgeon, to restore at least a semblance of the patient's former appearance. Lips were renewed, new noses built up, eyelids replaced, cavities in the palate filled in by flaps taken from the skin or scalp. The scar-tissues and flaps were kept pliant and adaptable by "spraying" with X-rays, which also served to depilate hair and to stimulate the healing process in both flaps and bone.

§ (31) X-ray Protection.—As already remarked, the X-rays possess in a marked degree the property of affecting animal tissue. As many of the earlier workers found to their cost, the effects of over-exposure to the rays may ultimately prove fatal. In view of the great public interest which the subject has excited, and the fact that complete safeguards are possible, the recent (1921) report is appended of a representative committee which reviewed the whole question in the light of present-day knowledge. The recommendations are many-sided and cover most of the points which experience has suggested both in X-ray and radium practice.

It may here be remarked that with exciting voltages in the region of 200,000, 3 mm. of lead reduces the intensity of X-rays over 10,000 times; 10 mm. of lead over 1 million times. Good lead-impregnated rubber sheet is ordinarily equivalent to about ⅓ or ¼ of the same thickness of lead sheet. For lead glass the corresponding figure is ⅓ to ⅙, although it is possible to obtain glass with a figure of ¼.

The dangers of scattered X-rays should be realised, particularly with very high exciting voltages. One of the best remedies is to enclose the bulb itself as completely as possible with adequate protective material. The report in question is as follows:

Introduction

The danger of over-exposure to X-rays and radium can be avoided by the provision of efficient protection and suitable working conditions.

The known effects on the operator to be guarded against are:

(1) Visible injuries to the superficial tissues which may result in permanent damage.

(2) Derangements of internal organs and changes in the blood. These are especially important, as their early manifestation is often unrecognised.

General Recommendations

It is the duty of those in charge of X-ray and Radium Departments to ensure efficient protection and suitable working conditions for the personnel.

The following precautions are recommended:

(1) Not more than seven working hours a day.

(2) Sundays and two half-days off duty each week, to be spent as much as possible out of doors.

(3) An annual holiday of one month or two separate fortnights.

Sisters and nurses, employed as whole-time workers in X-ray and Radium Departments, should not be called upon for any other hospital service.

Protective Measures

It cannot be insisted upon too strongly that a primary precaution in all X-ray work is to surround the X-ray bulb itself as completely as possible with adequate protective material, except for an aperture as small as possible for the work in hand.

The protective measures recommended are dealt with under the following sections:

I. X-rays for diagnostic purposes.
II. X-rays for superficial therapy.
III. X-rays for deep therapy.
IV. X-rays for industrial and research purposes.
V. Electrical precautions in X-ray Departments.
VI. Ventilation of X-ray Departments.
VII. Radium therapy.

It must be clearly understood that the protective measures recommended for these various purposes are not necessarily interchangeable; for instance, to use for deep therapy the measures intended for superficial therapy would probably subject the worker to serious injury.

I. *X-rays for Diagnostic Purposes*

(1) *Screen Examinations.*—(*a*) The X-ray bulb to be enclosed as completely as possible with protective material equivalent to not less than 2 mm. of lead. The material of the diaphragm to be equivalent to not less than 2 mm. of lead.

(*b*) The fluorescent screen to be fitted with lead glass equivalent to not less than 1 mm. of lead and to be large enough to cover the area irradiated when the diaphragm is opened to its widest. (Practical difficulties militate at present against the recommendation of a greater degree of protection.)

(*c*) A travelling protective screen, of material equivalent to not less than 2 mm. of lead, should be employed between the operator and the X-ray box.

(*d*) Protective gloves to be of lead rubber (or the like) equivalent to not less than ½ mm. of lead and to be lined with leather or other suitable material. (As practical difficulties militate at present against the recommendation of a greater degree of protection, all manipulations during screen examination should be reduced to a minimum.)

(*e*) The X-ray bulb to be used at as great a distance and emitting as little radiation as is consistent with the efficiency of the work in hand. Screen work to be as expeditious as possible.

(2) *Radiographic Examinations* ("*Overhead*" *Equipment*).—(*a*) The X-ray bulb to be enclosed as completely as possible with protective material equivalent to not less than 2 mm. of lead.

(*b*) The operator to stand behind a protective screen of material equivalent to not less than 2 mm. of lead.

II. *X-rays for Superficial Therapy*

It is difficult to define the line of demarcation between superficial and deep therapy.

For this reason it is recommended that, in the reorganisation of existing or the equipment of new X-ray departments, small cubicles should not be adopted, but that the precautionary measures suggested for deep therapy should be followed.

The definition of superficial therapy is considered to cover sets of apparatus giving a maximum of 100,000 volts (15 cm. spark-gap between points; 5 cm. spark-gap between spheres of diameter 5 cm.).

Cubicle System.—Where the cubicle system is already in existence it is recommended that

(1) The cubicle should be well lighted and ventilated, preferably provided with an exhaust electric fan in an outside wall or ventilated shaft. The controls of the X-ray apparatus to be outside the cubicle.

(2) The walls of the cubicle to be of material equivalent to not less than 2 mm. of lead. Windows to be of lead glass of equivalent thickness.

(3) The X-ray bulb to be enclosed as completely as possible with protective material equivalent to not less than 2 mm. of lead.

III. *X-rays for Deep Therapy*

This section refers to sets of apparatus giving voltages above 100,000.

(1) Small cubicles are not recommended.

(2) A large, lofty, well ventilated and lighted room to be provided.

(3) The X-ray bulb to be enclosed as completely as possible with protective material equivalent to not less than 3 mm. of lead.

(4) A separate enclosure to be provided for the operator, situated as far as possible from the X-ray bulb. All controls to be within this enclosure, the walls and windows of which to be of material equivalent to not less than 3 mm. of lead.

IV. *X-rays for Industrial and Research Purposes*

The preceding recommendations for voltages above and below 100,000 will probably apply to the majority of conditions under which X-rays are used for industrial and research purposes.

V. *Electrical Precautions in X-ray Departments*

The following recommendations are made:

(1) Wooden, cork, or rubber floors should be provided; existing concrete floors should be covered with one of the above materials.

(2) Stout metal tubes or rods should, wherever possible, be used instead of wires for conductors. Thickly insulated wire is preferable to bare wire. Slack or looped wires are to be avoided.

(3) All metal parts of the apparatus and room to be efficiently earthed.

(4) All main and supply switches should be very distinctly indicated. Wherever possible, double-pole switches should be used in preference to single-pole. Fuses no heavier than necessary for the purpose in hand should be used. Unemployed leads to the high-tension generator should not be permitted.

VI. *Ventilation of X-ray Departments*

(1) It is strongly recommended that the X-ray Department should not be below the ground level.

(2) The importance of adequate ventilation in both operating and dark rooms is supreme. Artificial ventilation is recommended in most cases. With very high potentials coronal discharges are difficult to avoid, and these produce ozone and nitrous fumes, both of which are prejudicial to the operator. Dark rooms should be capable of being readily opened up to sunshine and fresh air when not in use. The walls and ceilings of dark rooms are best painted some more cheerful hue than black.

VII. *Radium Therapy*

The following protective measures are recommended for the handling of quantities of radium up to one gram:

(1) In order to avoid injury to the fingers the radium, whether in the form of applicators of radium salt or in the form of emanation tubes, should be always manipulated with forceps or similar instruments, and it should be carried from place to place in long-handled boxes lined on all sides with 1 cm. of lead.

(2) In order to avoid the penetrating rays of radium all manipulations should be carried out as rapidly as possible, and the operator should not remain in the vicinity of radium for longer than is necessary.

The radium when not in use should be stored in an enclosure, the wall thickness of which should be equivalent to not less than 8 cm. of lead.

(3) In the handling of emanation all manipulations should, as far as possible, be carried out during its relatively inactive state. In manipulations where emanation is likely to come into direct contact with the fingers thin rubber gloves should be worn. The escape of emanation should be very carefully guarded against, and the room in which it is prepared should be provided with an exhaust electric fan.

Existing Facilities for ensuring Safety of Operators

The governing bodies of many institutions where radiological work is carried on may wish to have further guarantees of the general safety of the conditions under which their personnel work.

(1) Although the Committee believe that an adequate degree of safety would result if the recommendations now put forward were acted upon, they would point out that this is entirely dependent upon the loyal co-operation of the personnel in following the precautionary measures outlined for their benefit.

(2) The Committee would also point out that the National Physical Laboratory, Teddington, is prepared to carry out exact measurements upon X-ray protective materials and to arrange for periodic inspection of existing installations on the lines of the present recommendations.

(3) Further, in view of the varying susceptibilities of workers to radiation, the Committee recommend that wherever possible periodic tests, *e.g.* every three months, be made upon the blood of the personnel, so that any changes which occur may be recognised at an early stage. In the present state of our knowledge it is difficult to decide when small variations from the normal blood-count become significant.

VI. X-RAYS AND MATERIALS

In wellnigh every branch of industry the testing of materials has come to be of importance. With increasing knowledge and the stress of competition, a variety of testing methods have been evolved to ascertain quality and uniformity as determined by the several physical, chemical, and visual characteristics. Such tests are commonly conducted on samples which are selected to be as representative as possible. From the nature of things the value of the results is limited, and the engineer in particular is ever on the look-out for opportunities for further insight into the materials he employs.

The employment of X-rays in the examination of materials lies at present in two main directions:

(1) X-ray crystallography or the study of crystal structure.

(2) Radiography or X-ray shadow photography.

§ (32) X-RAY CRYSTALLOGRAPHY.—We can only refer to the great potentialities of the results of X-ray analysis as applied to crystal structure.[1] It is a matter of great satisfaction to Englishmen to know how much the subject owes to Sir Wm. Bragg and his son, whose published work on the subject is of the highest fascination and importance. Several methods have been employed. In 1912 Laue at Munich sent a heterogeneous beam of X-rays through a thin crystal and photographically showed that a diffraction pattern was produced. The Braggs followed with the X-ray spectrometer in which monochromatic X-rays are reflected from the several faces of a crystal, and by that means proceeded to disclose the atomic architecture of a large number of crystals.

The practical possibilities were greatly enlarged when Debye and Scherrer (at Zürich) and Hull (at the G.E.C. Research Laboratory, Schenectady) showed that large crystals were not essential, but that the method could be applied to an aggregate of finely powdered crystalline material, provided the orientation of the crystals were sufficiently random. This was a big step forward, for it enables the crystalline structure of a body to be examined even when the individual crystals are microscopic or ultra-microscopic in size. We now know that almost every solid substance betrays crystalline structure, and it would seem that the various physical properties—elasticity, hardness, melting points, etc.—are all manifestations of the various atomic forces which reveal themselves in the crystalline form. The very formation of solids may be merely an outward and visible sign of crystallisation, and a definition of a "solid" may be so derived which is, at any rate, as adequate as

[1] See also "Crystallography," § (11).

others which have been framed. Not only the growth but the decay, the change-points, etc., can all be followed and watched without harming the body in any way.

We have thus a new tool of research which, although at present rather delicate and tentative in application, would seem to offer boundless possibilities. The metallurgist, to whom crystalline formation means so much, need no longer have to content himself with inferring from their external forms what the internal structure of the crystals in his metals and alloys may be. He may also find that the method will throw light on the fundamental nature of the effect of heat treatment, tempering, rolling and ageing on steels and other crystalline metals and alloys. It has been shown that amorphous carbon really consists of minute graphite crystals; colloidal gold and silver are made up of minute, yet perfect, crystals, so small that they contain only a few score atoms. Even the particles "sputtered" from a cathode in a discharge tube are possible of examination and are found to be crystalline.

These are but a few of many examples. There is a great opportunity for the metallurgist and physicist to get together. At present the main difficulties are those of technique. Monochromatic X-rays have to be used, and as these can be obtained only of feeble intensity, protracted exposures have hitherto been necessary, though these can now be greatly shortened by the use of more sensitive plates.

§ (33) Industrial Radiography.—As was anticipated by Röntgen and others, when the art of radiography had sufficiently advanced in medicine it extended its scope to industry. As already remarked, the method of X-ray inspection has the advantage of not injuring a body in any way. Furthermore, it provides in many cases the only means of detecting concealed defects in a material, or of scrutinising in a structure the accuracy of assembly of component parts which are hidden from view.

The development of industrial radiology has been bound up with that of the Coolidge tube, and, both during and since the war, the X-rays have been applied to a variety of branches of industry. As already explained, the method depends on receiving the shadow of the object on a fluorescent screen or photographic plate, and it should be made clear at the outset that a radiograph shows only the gross structure of a material and gives no information as to the crystalline or microscopic structure from point to point.

While the general technique is much the same as in medicine, mention should be made of one of the chief experimental precautions in the X-ray photography of metals. Even in medical radiography the experienced worker is well aware of the effect of the scattered radiation which is generated whenever a beam of X-rays strikes any particle of matter. Such scattered radiation, if allowed to reach the photographic plate, tends to fog the main image. The various surfaces of the bodies encountered are the chief offenders, and even the air contributes its quotum.

The effect is especially marked with metallic objects which require relatively long exposures; worthless results will be obtained in the absence of suitable precautions. These consist in enveloping the photographic plate, back and front, with sheet lead (preferably with an inner lining of aluminium), a hole being left no bigger than necessary for the reception of the direct image of the object. If the object is continuous and flat there is no difficulty, for it can be brought into close contact with the plate. If, however, the body is irregular in contour, it may conveniently be cemented with paraffin wax to the bottom of a cardboard or aluminium tray, and mercury, fine lead shot, or the like poured round it. Wax filling is also necessary, both to fill up any pockets or cavities and to prevent the mercury or shot from straying into the path of the projected image.

Considerable gain may result from the use of the Bucky grid between the object and the plate. This consists of a rectangular metal grid, the faces being spherical in contour and the dividing cell-walls of the grid everywhere radial. The grid, while allowing direct X-rays from the focus to pass, kills the majority of the scattered radiation. The grid is kept in slight motion to prevent its being registered on the photograph. Still greater freedom from the effects of secondary radiation may be obtained by using specially sensitive plates and so shortening the exposure.

Naturally the orientation of the object with reference to the beam of X-rays may make or mar a radiograph. Distortion may be reduced by avoiding undue obliquity of the rays, and to this end it is wise to keep the distance between the object and bulb as great as is expedient. For good definition the rays should be stopped down as much as possible.

The present practicable depths which can be penetrated in various materials are:

4 to 5 mm. of lead.
12 mm. of tin.
7·5 cm. of steel (carbon) or iron.
10 to 15 cm. of aluminium and its alloys.
30 to 40 cm. of wood.

The limiting factor in practice is the exposure, which hitherto has been very protracted with the greater thicknesses. However, with the latest type of X-ray plate the exposures are greatly reduced, and 1 inch of steel, for example, now requires an exposure of rather less than a minute, using a voltage of about 130,000 and a few milliamperes through the tube.

Within the above limits we can, with considerable delicacy, hunt out anything which is so disposed as to cast a measurable variation in the shadow, provided the body is not too complicated in design to render the shadow too confusing to interpret. The method is surprisingly sensitive; for example, tool-marks and fine mould-marks often show up in a radiograph. The opacity is merely a measure of the number and weight of the atoms encountered, and so different qualities of a metal possessing different densities display different intensities in a radiograph; for example, a wrought rivet in a casting of the same metal shows a darker image. For the same reason, equal thicknesses of carbon, nickel, and tungsten steels differ markedly in transparency, a property which has been turned to account.

FIG. 19.—Defective Weld in Steel Plates.

Electric and oxy-acetylene welding have come into great prominence during and since the war; an indifferent welder can turn out what appears on the surface to be an excellent weld, but is quite an unreliable job notwithstanding. There appears to be no adequate mechanical test for a weld, and in any event any such test, whether mechanical or microscopical, destroys the weld, good or bad. The X-rays promise to be of great use in this connection. If the component parts are not actually fused together a narrow dividing line comes out on the plate. Blisters and blowholes show up as light spots. X-ray photography of welds up to 1 inch thick is now quick, easy, and certain: with modern equipment, lengths up to 2 feet can be taken at once, the exposure being a fraction of a minute. The amount of detail revealed is extraordinary, and the process compares favourably with that of photomicrography, which is only very local in its test and, as already remarked, involves the destruction of the weld. The X-ray method has proved to be a somewhat severe critic of present-day welds as commonly carried out (*Fig.* 19).

Hidden cracks in a metal, which are a bugbear to metallurgists, can often be detected, though if they are very fine or tortuous (hair cracks) the method is rarely suitable. Such cracks are sometimes the sequel to "pipes" or blowholes in the ingot, and it is easier to detect them in the ingot than after working.

In the case of alloys, the uneven distribution of any component results in a "patchy" or streaky radiograph. X-ray examination will often diagnose defective soldering or brazing, the substitution of one metal by another, hidden stopping or pinning, and so on. The method has also found application in detecting hidden corrosion (as in gas cylinders, in ferro-concrete, and the armouring of cables), in scrutinising steel turbine discs for segregations, etc., and so on.

Naturally enough, the X-rays found a great opening, during the war, in the manufacture of explosives and related devices. In some instances, *e.g.* the correct filling of liquid-gas grenades, the examination of opaque cordite, the interior detail of detonators, Stokes igniters, vent-sealing tubes and other pyrotechnic stores, no other method of inspection was possible. The X-rays also proved of value in examining enemy ammunition of unknown design, where, for reasons of safety, it was desirable to ascertain the internal construction before opening up. They have also proved useful in checking the contents of packing boxes. Most of this work was carried out by the Research Department at the Royal Arsenal, Woolwich, and the Editor is indebted to the Department for *Figs.* 20-23, illustrating some of the results arrived at.

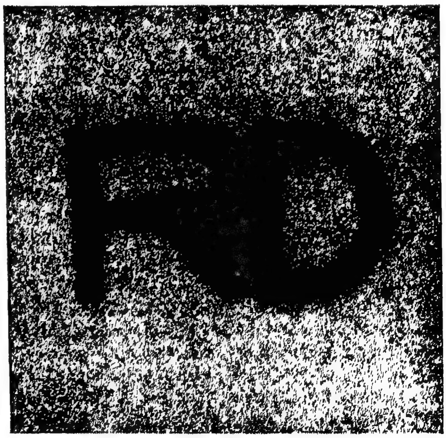


FIG. 20.—Radiograph of Lead Letters RD taken through a Steel Slab 3″ in thickness. (Radiological Laboratory, Research Department, Royal Arsenal, Woolwich.)

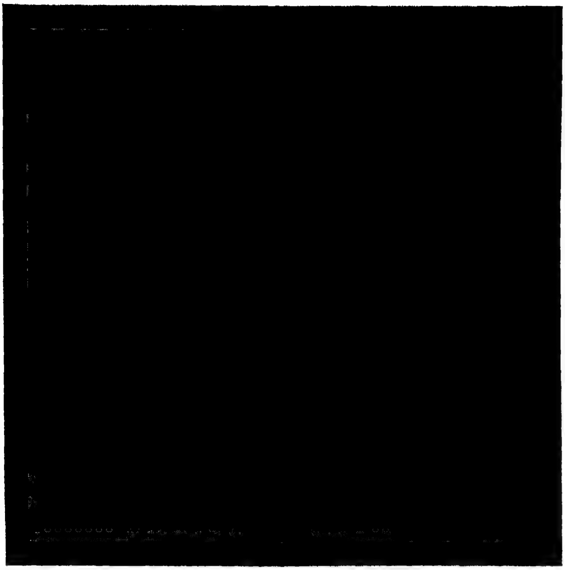


FIG. 21.—Radiograph showing Hidden Cracks in 1″ Boiler Plate. (Radiological Laboratory, Research Department, Royal Arsenal, Woolwich.)

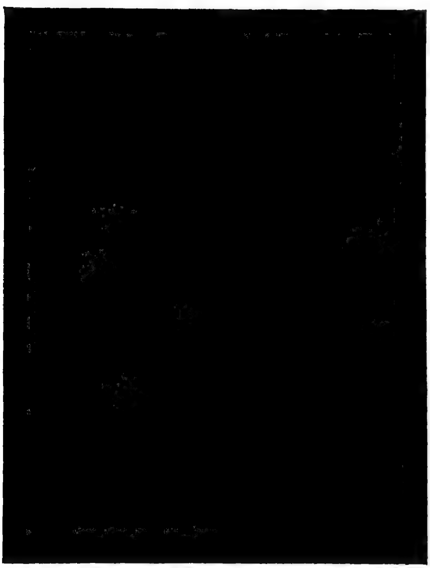


FIG. 22.—Radiograph showing Faulty Weld in Steel Plate ½″ in thickness. (Radiological Laboratory, Research Department, Royal Arsenal, Woolwich.)

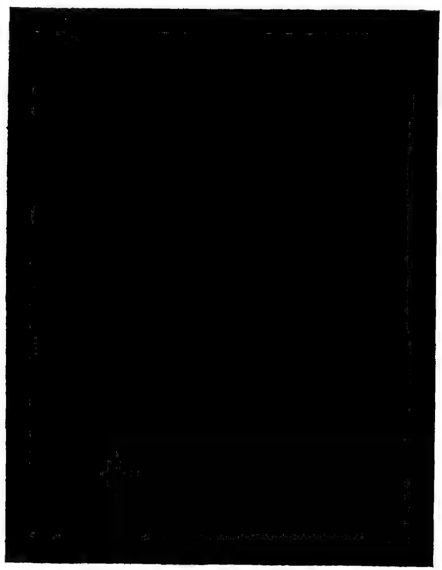


FIG. 23.—Radiograph showing Hidden Crack in Welded Steel Plate ½″ in thickness. (Radiological Laboratory, Research Department, Royal Arsenal, Woolwich.)

In the case of timber, the different varieties absorb X-rays to different degrees. Peculiarities in the structure and path of the fibres (such as the contortions which produce "figure") are easily discerned. The denser heart wood is differentiated from the sap wood, the summer and spring growths of the annual rings are readily identified (*Fig.* 24),

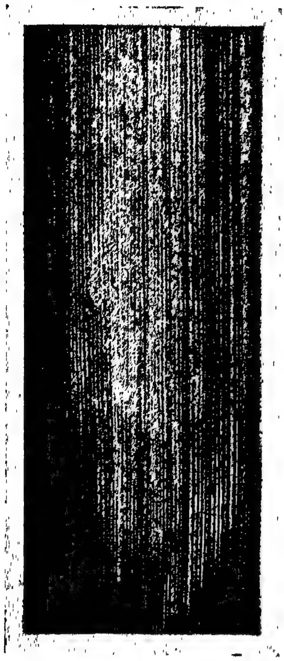


FIG. 24.—Perfect Specimen of Aeroplane Spruce.

and defects such as knots or grub-holes show up with astonishing clearness (*Fig.* 25).

A method of utilising the X-rays to examine

FIG. 25.—Concealed Knots and Grub-hole in Laminated Spar.

the wooden parts of aircraft was developed on behalf of the Air Ministry during the war. At a time when the submarine was seriously endangering the country's supplies of high-grade timber from Canada and the States, designs for building up aeroplane parts from smaller timber were developed, using laminated or "box" structures. The workmanship

FIG. 26.—Defective Shaping of End Block of "Box" Spar; also Block Split by Screws.

required has to be of the finest, and much of it is hidden of necessity, but the inspector now has a powerful ally in the X-rays, which unerringly reveal hidden faults such as knots, large resin-pockets, defective gluing, and poor workmanship (*Figs.* 26, 27, 28). Wood is

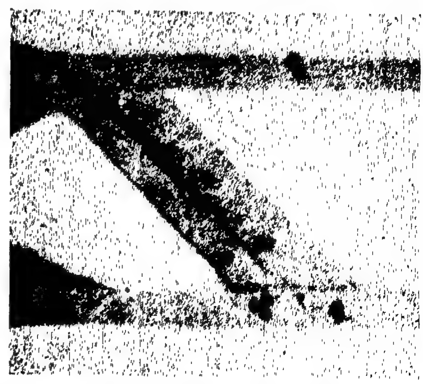


FIG. 27.—Aeroplane (Box) Spar showing forbidden Joint in Plywood Side.

very transparent to X-rays, and thicknesses up to 18 inches or more can be dealt with, screen examination being possible in most cases. The method is also useful for watching the behaviour of the various hidden

members and joints of a composite wooden structure while it is being subjected to test.

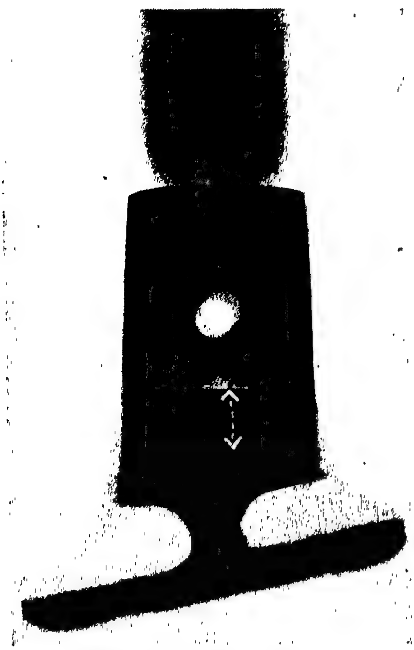


FIG. 28.—Defective "Bottoming" of Aeroplane Strut in Aluminium Socket.

X-rays are also being turned to account by the tyre manufacturer in his efforts to improve the union between the rubber and the Egyptian cotton fabric. In the manufacture of golf balls, fine rubber tape is wound on a round core either of soft rubber or liquid. If care is not taken the core is distorted, becoming either roughly ellipsoidal or even dumb-bell shaped. The resulting ball is defective from the point of view of accurate flight, but such balls can be readily sorted out by the help of the X-rays. The method is now in extensive use, no other being readily available (*Fig.* 29). The Post Office has long used the rays for testing the amount of mineral matter in gutta-percha.

The help of the X-rays has also been effectively sought by the manufacturer of carbon and graphite brushes and electrodes, to reveal mineral matter and internal cracks and flaws. The makers of electrical insulators —ebonite, built-up mica, fibre, paper, etc.— find the method invaluable for detecting the presence of metallic particles, often from the steel rollers used in the preparation of the material.

The manufacture of optical glass became a key industry during the war, as hitherto we had relied wholly on Germany for our supplies. One of the greatest troubles which was encountered was the destructive action of the molten glass on the fire-clay pot, in which the components were fused. It was found that the effect was caused by the presence of iron and other impurities in the clay. Recourse was had to the X-rays, and it was found that on examining the pots before they were fired, those containing prejudicial foreign matter could readily be sorted out. In this way much expense can be saved. The "melt" of optical glass can also be examined for inclusions before working.

X-ray photographs are useful for displaying the arrangement of concealed wiring, for example, when embedded in the interior of insulating panels or in radio apparatus. In much the same way, during the war, the X-rays were useful in scrutinising the wiring within the leather of aeroplane pilots' electrically heated clothing.

A similar field of work which the X-rays have found is the examination of the interior of moulded articles, for example, the distributors of magnetos. During construction the insulation is moulded round the metal work, and subsequently machined. If, during the machining, blowholes are met with, the entire distributor has to be rejected.

Among the miscellaneous uses of the X-rays we can only make mention of the examination of oysters for pearls; the differentiation of lead

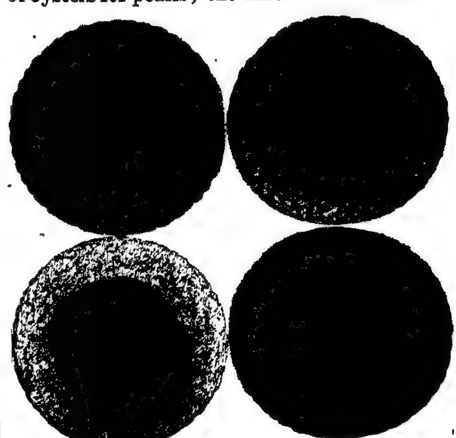


FIG. 29.—Golf Balls showing Unsymmetrical Cores.

glass jewels from the more transparent genuine gems; the scrutiny of artificial teeth; the detection of contraband by the customs officials; the sorting of fresh from stale eggs; the detection of heavy elements in minerals, of metal particles in chocolate, of weevils in grain, of mineral adulterants in certain powdered drugs (*e.g.* asafoetida), and of moths in tobacco for cigars.

In quite a different direction, an enterprising American shoe store has installed a screen outfit, so that the potential customer can see his "footigraph" and satisfy himself visually whether or not the shoe he is trying on is a good fit.

The application of the cinematograph principle to X-ray photography offers wide possibilities.

We can only refer to the more academic applications of the X-rays by the conchologist to examine the interior of shells and fossils, without in any way spoiling a rare specimen. These have valuable educational possibilities.

The use of the X-rays for revealing the interior of plant life is comparatively recent. Considerable differences exist in the mineral content and density, and hence the transparency of the different parts of a plant—root, stem, leaf, flower, fruit, seed, etc.—and thus it happens that even the most delicate structures of plants can be laid bare without tearing the plant to pieces in order to study it. Microscopic detail is of course not revealed. Long-waved X-rays are required for such work.

§ (34) X-rays and Old Masters.—The first artistic oil-painting of which there is any record was executed in the year 1399 by Hubert van Eyck, a Dutchman, and from then up to the early Italian and Flemish school, painters had possibly only eight or nine pigments, mostly mineral in origin. To-day there are over 200 in use, many of them vegetable or coal-tar in origin.

As is well known, the imitating of valuable pictures has always enjoyed a great vogue, and there are thousands of spurious paintings in existence—copies of both late and modern masters—which have been passed as genuine and sold for outstanding amounts. For example, so far as is known, Rembrandt painted some 700 pictures, yet Maximilian Toch estimates there are fully 4000-5000 in existence, all of which are regarded as genuine and have commanded great prices. Again it would have been absolutely impossible for any human being to have painted all the Rubenses that there are in existence. The remark is probably true of every great painter.

There are various scientific methods of determining the originality and age of paintings. Photomicrography is of great help: for example, in the case of a panel of a picture 300 years old the protoplasm in the cells of the wood has entirely dried out, a feature distinct from a modern panel. Chemical analysis of tiny detached fragments often throws light on the subject; for instance zinc white (zinc oxide) was not known 300 years ago, and the Flemish painters used flake white (white lead). Again bitumen, at first transparent, gradually becomes opaque and insoluble with the passage of time.

But not only pictures, but all works of art, are imitated in the same way. Furniture, pottery, bronzes, old weapons and brass work are so completely simulated that experts are frequently baffled. Sheraton furniture is a familiar example. We know that Sheraton had a little shop and did most of his work himself with only occasional help from a few expert artisans. The amount of Sheraton furniture in existence would indicate that Sheraton had a factory of several acres employing a thousand men who were lifelong supporters of mass production.

It would appear that the X-rays may usefully be called in, in certain cases, as a supplementary method of scrutiny for the expert. A start has been made with pictures, as we shall now proceed to show.

In any picture we have to consider three media, (1) the surface which is painted on—usually canvas or wood, though paper, porcelain, or other materials may be used; (2) the priming or sizing—nowadays almost always white lead, though formerly carbonate of lime and glue were employed; (3) the actual pigments.

Both wood and canvas are very transparent to the X-rays, though different kinds of canvas vary a good deal. The white lead primer is much more opaque than carbonate of lime, and the former, moreover, penetrates much farther into the interstices of the canvas. This in itself is sufficient to show a marked difference under the X-rays between modern and older pictures.

As to pigments, they vary greatly in X-ray opacity from the opaque salts of lead, zinc, and mercury to the transparent aniline derivatives and bitumen. Both modern and ancient whites are usually opaque, most of the blacks (new or old) are transparent, and modern reds are more transparent than the old reds. But, as already remarked, most of the earliest pigments are mineral in origin and opaque.

In a modern picture the sizing is very commonly more opaque than the pigments, and X-ray examination is, for that reason, usually inconclusive. But fortunately in the pictures of the old masters the reverse conditions hold, and thus it is that with a little experience the X-rays can be employed most usefully as a means of identifying a modern fake, or detecting alterations to an old picture. It is a practical certainty that, however skilfully the process has been carried out, the several materials used—whether canvas, priming, or pigment—will differ from those in the original painting and will, in consequence, be differentiated in the radiograph.

Notable work on this subject has been carried out by Dr. Heilbron of Amsterdam and, more recently, by Dr. Chéron of Paris.

Among the sixteenth century paintings examined by the former was the "Crucifixion" by Cornelis Engelbrechtsen, which contained in the right foreground the portrait of a woman which it was suspected was that of a former "donatrice," who (after a fashion not unknown in those days) had thus sought to perpetuate her association with the picture. A radiograph of the painting showed many "restorations," especially on the right half, and beneath the portrait of the donatrice was revealed the picture of a monk in surplice and stole, the head being smaller than that of the over-painted lady. The evidence was so clear that the picture was sent to be restored at the Rijks-museum in Amsterdam, the result being to bring to light once more the monk who had been hidden for 400 years.

Among the other paintings examined by Heilbron was a panel of the "Madonna" by Geertgen van St. Jans (*c.* 1500) which had always excited comment because of the apparently stiff and unnatural position of the arms. The radiograph showed that the presence of the Child in the arms of the Madonna fully explained their attitude. St. Jans is known to have painted his children disproportionately small, and the presumption is that this defect was the cause of some former owner having the Child painted out.

Other examples of Heilbron's work include a panel by De Meester van Allmaar, where the portrait of a lady (again supposed to be the donatrice) is found to be painted over the original figures. There is some chance that the panel will be restored to its original state. A radiograph of a panel by van Dyk, representing a waterfall, a knight with a horse, dogs, etc., shows that the artist originally painted a much bigger waterfall, the current of water appearing to pass through the animals. We are led to infer that the painting is an original and not a copy, for only in the case of the original can we trace such alterations in the ideas of the artist.

Dr. Chéron X-rayed a Flemish panel attributed to van Ostade and showing a party of country dancers and revellers. The radiograph revealed only a farmyard scene containing peacocks, ducks, and chickens. The supposed van Ostade is almost certainly modern, since practically all its colours are transparent to the rays. The farmyard picture is apparently old, since the sizing is not opaque.

Another picture of the French school of the fifteenth century which was examined by Chéron was that of the Royal Infant at Prayer hanging in the Louvre. The black background was found to mask a badly deteriorated original background—confirming documentary evidence to that effect.

The X-rays may find another field in the examination of palimpsests and ancient manuscripts which, hitherto regarded as carrying only their face value, may bear under the trivial inscriptions of mediaeval times older matter of priceless worth. Again it is well known that before millboard came into general use for book covers (about the middle of the sixteenth century), binders were accustomed to make them up from such loose pages as came to hand. Many discoveries of rare and valuable MSS. have been made when the bindings of old volumes have happened to fall to pieces. The X-rays may have useful application here.

As regards antique furniture and the like, it is not improbable that examination of constructional or other detail, which cannot otherwise be viewed except by destroying the article, would suffice to reveal in a fake craftsmanship out of tune with the reputed period.

§ (35) Future Developments of Industrial Radiology.—Our ideal should be to make the taking of an X-ray photograph as easy as that of light. The present limitations of radiometallography are largely those prescribed by equipment and technique. Considerable improvements will have to come if the subject is to extend its scope and become an attractive commercial proposition in heavy engineering. If great thicknesses are to be tackled, means will have to be found so that exposures are not intolerably long. There appear to be two means to this end—(*a*) by using much heavier X-ray outputs, at much higher voltages, or (*b*) by using much more sensitive screens, plates, or other detectors.

We have already considered the probable developments of the high-potential generator, and, as we should anticipate, all experience agrees in demanding higher and higher voltages for work with metals. The ordinary Coolidge tube will, however, take no more than 180,000 volts, preferably less. This can be increased to 300,000 by lengthening the arms of the tube and completely immersing it in oil. If there is a demand for it, the electrical engineer will doubtless overcome the difficulties in the way of supplying half a million or more volts. Such transformers have already been made for other purposes, but their bulk, weight, and cost are formidable. For example, a single-phase transformer giving a peak voltage of one million occupies a floor space of 13 ft. × 8 ft., is 15 ft. high, has terminals 28 ft. high, weighs 20 tons, and costs about £10,000. With such voltages both transformer and tube will doubtless be contained in a common oil tank, thus reducing the danger and the considerable losses by brush discharge.

Heavier discharges will demand more elaborate cooling arrangements, and probably glass X-ray tubes will not stand up to the work. We may have to turn to metal tubes radically different in design, capable of

absorbing 50 h.p. or more. Furthermore, we shall have to improve the deplorable efficiency of the whole outfit.

It may be mentioned that some of the γ-rays of radium are far more penetrating than the hardest X-rays we can produce at present (being equivalent to X-rays excited by about two million volts), but unfortunately the intensity is so weak (not more than a few per cent of that from a good bulb) that exposures are intolerably protracted.

As regards fluorescent screens and photographic plates, great improvements are called for. No screen at present available is sensitive enough for thicknesses exceeding about $\frac{1}{4}$ inch steel, and only then with difficulty. Photography must be resorted to in such cases, and the time taken over the process may then become prohibitive, at any rate for routine "mass inspection."

A photographic plate registers only about 1 per cent of the X-rays passing through it. Progress has mainly consisted in thickening the emulsion or richly loading it either with more silver or with heavier metals. Exposures may be shortened either by backing up the emulsion with a sheet of a heavy metal, such as lead, or more appreciably by the use of an intensifying screen, containing a fluorescing salt, such as calcium tungstate. All X-ray plates are much more sensitive to visible rays than to X-rays, but such screens, which are more efficacious with "hard" rays than "soft," are apt to impair the detail in certain classes of work owing to "grain." It is important to have the closest contact between the screen and the emulsion. This is secured in the new "Impex" plate, in which the fluorescing salt is contained in a superimposed gelatine film which is dissolved off before the plate is developed. Such plates reduce the exposure as much as thirty times with very hard rays, though the efficiency is much less with longer waves.

Another real advance in X-ray photography has proved to be the duplitized film, *i.e.* a film coated with emulsion on both sides of the celluloid. A "pile" of several of these sandwiched with thin fluorescent screens, gives a very sensitive detector.

The ionisation method of detecting the X-rays offers great promise, for it can be made more sensitive than any photographic method at present available. An explorer built on these lines and of convenient design would have corresponding advantages.

To conclude, the subject of industrial radiology is young and, although progress has been rapid, we must in all fairness be careful not to claim too much for it. From such experience as we have had, it does appear, however, that the method is settling down to be a valuable laboratory tool, supplementing those which are already available for testing materials. Incidentally the existence of the method is not without its moral effect on the personnel, as regards standard of workmanship.

G. W. C. K.

Radiometer: an instrument for investigation of the infra-red spectrum; a modification of the instrument, devised by Crookes, in which mica vanes, accurately mounted on a central spindle *in vacuo*, rotate when placed in the path of radiant energy. See "Wave-lengths, The Measurement of," § (7).

Radiometry: the measurement of radiant energy or radiant power. See "Spectrophotometry," § (2).

Radio-micrometer: an instrument for investigation of the infra-red spectrum; a type of galvanometer in which the current is generated by radiation falling on the junction of a thermocouple. See "Wave-lengths, The Measurement of," § (7).

Radiotherapy: X-ray treatment of malignant disease. See "Radiology," § (30).

RADIUM

§ (1) Discovery.—Radium was discovered by M. and Mme Curie shortly after the property of radioactivity of uranium compounds had been discovered by Henri Becquerel in 1896. From observations of the activity of different compounds of uranium, Mme Curie had concluded that the radioactivity of uranium was an atomic property; that is, the total radiation emitted by a compound was proportional to the amount of uranium present, irrespective of the other inactive elements with which it was held in chemical combination. When, however, a number of uranium minerals were examined it was found that some specimens of pitchblende, notably those from Austrian mines, showed much greater activity than the metal uranium itself. It was suspected that this abnormal activity was due to the presence of small quantities of an unknown element or elements of activity greatly exceeding that of uranium. With the object of testing this conclusion, M. and Mme Curie undertook careful analyses of some of the uranium minerals, and their work resulted in the discovery of two new radioactive elements, namely polonium and radium, the activity of the latter being several million times that of an equal weight of uranium.[1]

§ (2) Occurrence.—The two principal commercial ores of radium are pitchblende and carnotite. Pitchblende has no definite composition and is a very complex mineral.

[1] M. and Mme Curie and G. Bemont, *Comptes Rendus*, 1898, cxxvii. 1215.

It contains in varying quantities nearly all the known metals, but it is rich in the oxides of uranium. Carnotite has a more definite composition, being a potassium uranyl vanadate (K_2O, $2UO_3$, V_2O_5, $3H_2O$) containing small quantities of barium and calcium. Of less importance are autunite, a hydrated calcium uranium phosphate [$Ca(UO_2)_2(PO_4)_28H_2O$] and torbernite, a hydrated copper uranium phosphate [$Cu(UO_2)_2P_2O_8$, $8H_2O$].

Pitchblende deposits are found at St. Joachimsthal, Bohemia; at Johanngeorgenstadt, Saxony; in Cornwall, England; in North Carolina and Connecticut, U.S.A. The mines of St. Joachimsthal have been worked for the last twenty-five years for uranium; previously they were worked for bismuth and cobalt, and before that for silver. The veins show that deposition occurred in three periods; the cobalt and nickel deposited first, then the uranium, and afterwards the silver. Dolomite spar is always present and generally has a white or yellowish-white colour, but changes to brownish red where pitchblende begins to appear, and is a dirty grey where it is actually in contact with the ore. The veins in the Saxony mines resemble those at Joachimsthal, the pitchblende occurring in the spar in pieces 2 to 3 in. in diameter. In Cornwall also the mineral pitchblende is associated with nickel and cobalt veins, although here only part of the veins are highly argentiferous. In the United States there are five mines that have produced pitchblende in quantity, and of these the Kirk mine is probably the most important. During the last twelve years about 20 tons of ore with an average content of 35 per cent U_3O_8, and over 100 tons with a content of 3 to 4 per cent U_3O_8, have been mined. The other mines produce lower grade ores in some quantity. Pitchblende has been found in small quantities, but of very high quality, in East Africa, and recently considerable amounts of an exceedingly rich uraninite have been found in India.

Carnotite deposits are found mainly in Dolores, San Miguel and Montrose counties, Colorado, and Utah, and extend over a belt about 60 miles long by 20 miles wide. The most usual ore is a sandstone so impregnated with yellow carnotite that the colour is decidedly noticeable and contains pockets of brown sandclay. The deposits are invariably in pockets, many of which, however, are of considerable size. A survey of the carnotite fields, carried out by the Bureau of Mines in 1912, revealed the fact that the carnotite deposits of Colorado and Utah constituted by far the largest source of radium-bearing ores in the world.[1] It is estimated that these deposits should be capable of yielding at least 500 grammes of radium element. During the period from 1913 to 1919 the Standard Chemical Company, Pittsburg, Pa., alone produced 50·8 grammes of radium element from this ore.[2] Carnotite is also found at Olary, South Australia, but this variety, being mixed with ilmenite, is very different from the American carnotite. Autunite is found in commercial quantities in Portugal in the district between Guarda and Sabugal. It is occasionally found in very pure condition, but for the most part as a very low grade ore, bearing from 0·5 to 1 per cent of U_3O_8. It is also found together with torbernite near Farnia in South Australia.

All common rocks and minerals of the earth's crust contain minute amounts of radium (of the order of 10^{-12} gramme of radium per gramme of rock). The atmosphere also contains radium in the form of emanation.

§ (3) Extraction and Separation of Radium from Radioactive Ores.—The chief ore employed in the preparation of radium in the early stages of radioactivity was pitchblende. The following account of the extraction of radium from pitchblende is given by Mme Curie.[3] The crushed ore is roasted with sodium carbonate, and the resulting material washed first with warm water and then with dilute sulphuric acid. The solution contains the uranium, whilst the insoluble residue contains all the radium. This residue chiefly contains the sulphates of lead and calcium, silica, alumina, and ferric oxide. In addition, nearly all the metals are found in greater or smaller amount (copper, bismuth, zinc, cobalt, manganese, nickel, vanadium, antimony, thallium, rare earths, niobium, tantalum, arsenic, barium, etc.). Radium is found in this mixture as sulphate and is the least soluble sulphate in it. In order to dissolve it, it is necessary to remove the sulphuric acid as far as possible. To do this, the residue is boiled with concentrated soda solution. The sulphuric acid combined with the lead, aluminium, and calcium passes into solution as sulphate of sodium, which is removed by repeated washing with water. The alkaline solution removes at the same time lead, silicon, and aluminium. The insoluble portion is well washed and treated with ordinary hydrochloric acid, which completely disintegrates the material and dissolves most of it. The radium remains in the insoluble portion. This is well washed with water and again treated with boiling concentrated solution of sodium carbonate. This operation completes the transformation of the sulphates of barium and radium into

[1] R. B. Moore and Karl L. Kithil, U.S.A. Bureau of Mines, *Bull.*, 1913, No. 70.

[2] *Science*, 1919, xlix. 227.

[3] Translation of Thesis for Doctorate, *Chem. News*, 1903, p. 134.

carbonates. The solid is filtered off, washed with sodium carbonate solution and then with water, and dissolved in HCl quite free from H_2SO_4. The filtered chloride solution is treated with sulphuric acid to precipitate radium and barium sulphates which are contaminated with traces of lead, iron, and calcium sulphates. The crude sulphates thus obtained have an activity from thirty to sixty times as great as that of metallic uranium. To purify the sulphates they are boiled with sodium carbonate and transformed into the chlorides. The solution is treated with sulphuretted hydrogen, which gives a quantity of active sulphides containing polonium. This precipitate is removed by filtering, and the remaining solution oxidised by means of chlorine, and precipitated with pure ammonia; the precipitate which contains actinium is removed, and the filtered solution treated with sodium carbonate. The precipitated carbonates formed are washed and converted into chlorides. These chlorides are evaporated to dryness and washed with pure concentrated hydrochloric acid to remove traces of calcium chloride. The calcium chloride dissolves almost entirely, whilst the chlorides of barium and radium remain insoluble.

From 1 ton of Joachimsthal residues about 8 kilograms of a mixture of barium and radium chlorides are obtained, the activity of which is about sixty times that of metallic uranium. To extract pure radium chloride from barium chloride containing radium the mixture of the chlorides is subjected to fractional crystallisation in pure water, and then in the later stages in water to which hydrochloric acid has been added. The crystals of radium chloride which separate out are elongated needles, having exactly the same appearance as those of barium chloride. Both crystals show double refraction. Crystals of barium chloride containing traces of radium are colourless, but when the proportion of radium increases they have a yellow coloration after some hours verging on orange and sometimes a beautiful pink. The maximum coloration is obtained for a certain degree of radium present. This fact serves as a useful check on the progress of fractionation. Giesel pointed out that the separation of barium and radium by fractional crystallisation in water from a mixture of the bromides is more rapid and efficient than from that of the chlorides.[1] Another method of separating barium and radium is given by Marckwald.[2]

A method of extracting radium from pitchblende by fusing the ore with sodium sulphate was employed on a large scale by Haitigner, Luding, and Ulrich.[3] This method has also been used in Australia by Radcliffe in treating carnotite obtained at Olary, South Australia.

Fusion with sodium carbonate was employed in America on carnotite ores, but this method has the great disadvantage that it carries a large part, if not all, of the silica into solution. This adds greatly to the cost of the operation and tends to give radium-barium sulphates of a rather high degree of impurity which require special treatment. The method adopted at present by the Bureau of Mines, U.S.A.,[4] is based upon the fact that strong nitric acid dissolves radium-barium sulphates in considerable quantities, as well as the other soluble constituents of the ore. The use of strong nitric acid is advocated because the radium-barium sulphate can be precipitated at once in a remarkably pure form, and the nitric acid can be largely recovered in the form of sodium nitrate and used again. The radium usually exists in the ore in the proportion of 1 part in 200,000,000; a recovery of 90 per cent of the radium present is claimed for the nitric acid method. H. Schlundt[5] gives a method of extracting radium from carnotite ores with concentrated sulphuric acid, and has applied it with success to carnotite ores from Colorado and Utah. By boiling carnotite ores with concentrated sulphuric acid the barium and radium compounds present are converted into bisulphates, which remain in solution in an excess of the acid and may then be separated from the insoluble components by filtration, followed by washing the residue with concentrated sulphuric acid. From the acid liquors thus obtained the radium is recovered by diluting with water, whereby radium-barium sulphate is precipitated. Two types of carnotite ores were dealt with, a low-grade ore containing 1·66 per cent U_3O_8, 4·03 per cent V_2O_5, and 4·88 parts of Ra per billion, and a high-grade ore containing 14·39 per cent U_3O_8, 9·67 per cent V_2O_5, and 42·78 parts of Ra per billion. By using sulphuric acid of more than 78 per cent concentration, at least 90 per cent of the radium present in these ores may be recovered by the above method.

Plum[6] gives the following summary of the method employed by him in separating radioactive products from carnotite concentrates. One kilogramme of carnotite concentrates was boiled for several hours with 2 litres of a solution containing 400 grammes of anhydrous sodium carbonate. The filtrate yielded 88·1 grammes

[1] Giesel, *Ann. Chem. Phys.*, 1899 (ii.), lxix. 91; Schall, *Am. Chem. Soc. J.*, 1920, xliii. 889-896.

[2] *Ber.* 1904, xxxvii. 88.

[3] *Ber. über die Bearbeitung der Pechblend Rückstände K.K. Akad., Wissenschaft*, 1908, cxvii. 619.

[4] C. L. Parsons, R. B. Moore, S. C. Lind, and O. C. Schaefer, *J. of Ind. and Eng. Chem.*, viii. No. 1.

[5] *Journ. of Phys. Chemistry*, 1916, xx. 485.

[6] *Journ. Am. Chem. Soc.* xxxvii. 8.

of uranyl sodium carbonate, $UO_2CO_32Na_2CO_3$, which represented 88 per cent of the total uranyl in the samples. The filtrate still held 2 per cent of the uranium in solution. A second treatment of the ore with 300 grammes of sodium carbonate under the same conditions as above dissolved out only 2·8 per cent of the uranium. The residue was treated with 400 c.c. hydrochloric acid diluted with about a litre of water, and boiled for about 8 hours. The radium-barium sulphate separated from the filtrate weighed 7·36 grammes. The residue was next heated for a day with 200 c.c. of nitric acid diluted with about a litre of water. The radium-barium sulphate precipitated in this filtrate weighed 0·453 gramme. The percentage of radium in these combined sulphates was 89·8 per cent of the total amount in the ore. Reprecipitations of barium sulphate in the two acid filtrates carried down 2·7 per cent more of the radium. The lead sulphate separated from bismuth and containing radio-lead, weighed 0·49 gramme. The polonium precipitated with bismuth and then deposited on copper was 50·1 per cent of the total amount in the ore. The residue was finally treated with twice its weight of sulphuric acid after being diluted with about an equal weight of water and then heated until most of the sulphuric acid had escaped in fumes. The activity of the ionium found in this solution was 61 per cent of the total amount in the original carnotite. The residue left, after all extraction processes had been carried out, weighed 507 grammes and its activity showed that only 4·2 per cent of the radium still remained in it.

§ (4) Metallic Radium has been isolated by Mme Curie and Debierne [1] by the electrolysis of pure radium chloride. The radium amalgam so obtained was heated to about 700° C. in a current of hydrogen to volatilise the mercury. A white metal remained. Metallic radium melts at about 700° C. and the vapour given off rapidly attacks quartz. The metal turns black in air, possibly owing to the formation of a nitride; it chars paper and dissolves rapidly and completely in water and in dilute hydrochloric acid. It decomposes water with the evolution of hydrogen forming radium hydroxide solution.

§ (5) Atomic Weight.—Mme Curie made several determinations of the atomic weight of radium, by treating purified radium chloride with silver nitrate and estimating the amount of silver chloride obtained. The last determination made by her gave a value 226·4 for the atomic weight.[2] Employing the same method and about 1·35 grammes of purified radium chloride O. Honigschmid [3] obtained the value 225·95. A similar set of experiments with the bromide gave 225·96. A spectroscopic examination of the final bromide preparation used showed that it contained less than 0·002 per cent of barium. The value 226 is at present accepted as the atomic weight of radium.

§ (6) Spectrum.—Demarçay [4] showed that radium gave a well-marked and characteristic spectrum. Twelve lines were observed between $\lambda=5000\ \mu\mu$, and $\lambda=3500\ \mu\mu$, together with well-marked bands. The general aspect of the spectrum is that of the alkaline earths—these metals have well-marked line spectra with a number of bands. Later the spark spectrum of radium was examined by Runge,[5] Exner and Haschek,[6] Crookes,[7] Runge and Precht.[8] The chief lines observed, in order of decreasing intensity, were: 3814·6, 4682·4, 4340·8, 5729·2, 4826·1, 4436·5, 5813·8, 6200·6, 5958·4, 5660·8, 4533·3, 2813·8 $\mu\mu$. Radium salts give a fine carmine coloration to the Bunsen flame. The flame spectrum of radium was examined by Runge and Precht.[9] The following lines and bands showed up clearly: 4826, 6130–6330, 6329, 6349, 6530–6700, and 6653. A good test of the completion of the separation of barium and radium is the relative intensities of the lines 4533·3 (radium) and 4554·2 (barium). These two lines are of same intensity when only 0·6 per cent barium is present in the radium salt. An activity fifty times as great as that of metallic uranium is required to distinguish clearly the principal radium line (3814·6 $\mu\mu$) in the spark spectrum. With a sensitive electrometer the radioactivity of a substance, only one-hundredth of that of metallic uranium, can be detected, so that the ionisation method of detecting radium is much more sensitive than the spectroscopic method.

§ (7) Salts of Radium.—The salts of radium resemble the corresponding salts of barium. Radium sulphate is less soluble than the barium salt, the carbonate also is sparingly soluble. For a 50 per cent or weaker solution of sulphuric acid the solubility of radium sulphate is practically the same as that in water, namely $2{\cdot}0\times10^{-8}$ gm. $RaSO_4$ per c.c. at 25° C. This value is that predicted from comparison with the decreasing solubilities of Ca, Sr, and Ba sulphates. On increasing the concentration of H_2SO_4 above 65 per cent, a marked rise in solubility of $RaSO_4$ takes place; it is more than twelve times as soluble in 70 per cent as in 65 per cent acid.[10]

[1] *Comptes Rendus*, 1910, cli. 523.
[2] *Ibid.*, 1907, cxlv. 422.
[3] *Wien. Ber.*, 1911, cxx. 1617.

[4] *Comptes Rendus* 1898, cxxvii. 1218.
[5] *Ann. d. Phys.*, 1900, ii. 742; 1903, xii. 407.
[6] *Wien. Ber.*, 1901, cx. 964.
[7] *Roy. Soc. Proc.*, 1904, A, lxxii. 295.
[8] *Ann. d. Phys.*, 1904, xiv. 418.
[9] *Ibid.*, 1903, x. 655.
[10] Underwood and Whittemore, *Am. Chem. Soc. J.*, 1918, xl.

The bromide and chloride crystallise with two molecules of water: $RaCl_2 2H_2O$, $RaBr_2 2H_2O$, and these crystals are isomorphous with the corresponding barium salts. The radium halides are much less soluble than the barium halides. Several other salts, for instance nitrate, azoimide, cyanoplatinate, have been prepared. The radium salts when freshly made are white, but they afterwards become yellow and brown, particularly if the salts are impure. Solutions of radium salts have a blue luminescence and the salts are all luminous in the dark. The following table gives the percentages of radium element in some of the radium salts:

TABLE I

Salt.	Formula.	Percentage of Radium Element in Pure Radium Salt.
		per cent
Radium bromide crystalline	$RaBr_2 2H_2O$	53·6
Radium bromide anhydrous	$RaBr_2$	58·6
Radium chloride crystalline	$RaCl_2 2H_2O$	67·9
Radium chloride anhydrous	$RaCl_2$	76·1
Radium sulphate	$RaSO_4$	70·2
Radium carbonate	$RaCo_3$	79 0

§ (8) RATIO OF RADIUM TO URANIUM IN MINERALS.—In the early stages of radioactivity Rutherford and Soddy had suggested that radium was a disintegration product of one of the radioactive substances in pitchblende and, since radium was always found associated with uranium, it appeared probable that uranium was the primary source from which radium was derived. If this were the case the weight of pure radium per gramme of uranium in a mineral in radioactive equilibrium should be a definite constant. A large number of radioactive minerals were examined by different investigators with a view to testing this conclusion. The ratio of the amount of radium to uranium in the older minerals showed remarkable constancy; minerals of more recent formation, however, in which radioactive equilibrium had not been established, showed variation, and according to expectation the values of the ratios were always lower than in the older minerals. The weight of pure radium per gramme of uranium in a mineral in radioactive equilibrium was found by Rutherford and Boltwood[1] to be $3{\cdot}23 \times 10^{-7}$ gramme, so that 323 milligrammes of radium element are present per ton of uranium element approximately. This value of the "equilibrium ratio" has been confirmed by Lund and Whittemore[2] who examined samples of carnotite representing large quantities of the ore (a few hundred pounds to several tons). Samples from small quantities of ore—hand specimens up to a few pounds—tended to exhibit abnormal ratios; in one instance the ratio was as low as $2{\cdot}48 \times 10^{-7}$, in another it was as high as $4{\cdot}6 \times 10^{-7}$. These abnormal differences are put down to local variations which are equalised in large samples.

[1] *Ann. Journ. Sci.*, 1906, xxii. 1.
[2] Bur. of Mines, *Techn. Papers*, 1915, lxxxviii.

§ (9) ESTIMATION OF RADIUM.—The quantity of radium in a preparation is estimated in terms of the International Radium standard. This standard is a specially purified specimen of radium chloride containing 21·99 milligrammes of radium chloride prepared by Mme. Curie. It is kept at the Bureau International des Poids et Mesures, Sèvres, Paris. Copies of this standard have been supplied to the National Physical Laboratory, Teddington, and to the official testing institutions of other countries, who undertake the determination of the quantity of radium contained in preparations submitted to them.

Two general methods have been employed for determining quantitatively the amount of radium in a preparation, namely, the γ-ray method which is suitable for measuring quantities of radium varying from 0·1 milligramme to 1 gramme, and the emanation method which is suitable for measuring small quantities of the order of 10^{-6} milligramme.

(i.) *The γ-ray Method.*—This method of measurement depends upon the fact that the radium in a radium salt, hermetically sealed in an enclosure, is in equilibrium with its products of disintegration in about a month's time after the preparation of the salt, and emits a very penetrating γ-radiation, the intensity of which is proportional to the amount of radium present. The radium itself emits no γ-radiation, neither do the next three products which follow it, except radium B which gives off soft γ-radiation; the penetrating γ-rays of radium are entirely due to the fourth product radium C, and it is for this reason that it is necessary to enclose the salt in a hermetically sealed tube, so that equilibrium has been established before measurements are taken. It is only when equilibrium has been attained that the γ-radiation gives a measure of the radium present in the tube.

(a) *Rutherford's Direct Method.*—To measure the quantity of radium in a sample in terms of the standard the γ-ray activity of each preparation is measured under identical conditions by means of an electroscope and ionisation chamber. The apparatus employed is shown in *Fig.* 1. The rates of fall of the electroscope leaf after correcting for natural leak are directly proportional to the amounts of radium in the tubes and thus give a direct comparison

of their amounts. It is essential for accurate measurements that the electric field inside the ionisation chamber is strong enough to produce saturation. The most accurate comparisons are those made when the sample and standard contain about the same quantity of material; comparisons correct to about 1 per

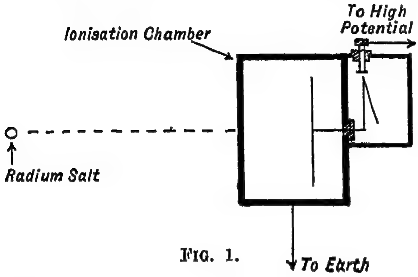


FIG. 1.

cent can be made, however, when one sample is ten times as strong as the other if a sufficient number of readings be taken. It is advisable to make comparisons at different distances from the ionisation chamber and also to place different thicknesses of absorbing screens in front of the chamber. The ratio of the activity of the two samples should remain constant during these modifications. Measurements should not be made with the samples too near the electroscope because a small displacement relative to the electroscope introduces a large error. It is essential to prevent radiation other than γ-radiation from entering the chamber. The latter is therefore made of lead at least 3 millimetres thick and the windows of the electroscope should be made of thick lead glass and be well guarded by lead shields. These precautions prevent the entry of β-radiation, but scattered γ-radiation from objects in the neighbourhood of the apparatus will always enter the apparatus, and for this reason such objects should not be moved during the course of a comparison. It is of advantage to enclose the radium in small tubes so that they act as point sources, also the walls of the tubes should not exceed 0·2 millimetre in thickness. If they do not exceed this thickness the correction due to absorption of the rays in the glass is negligible. For thicker walls, a correction will have to be applied, the absorption coefficient of the rays in glass being 0·10 per cm. It is convenient to have at hand a series of graduated standards, say 0·1, 1, 5, 10, 30 milligrammes radium element for the comparisons. The error introduced on account of lack of saturation may be serious when quantities of radium of very different activities are compared by the above method. The graduated standards would be useful in this connection, because they could be combined together to give an amount of radium approximately equal to that in the tube to be tested.

(*b*) *Rutherford and Chadwick's Method.*—Another delicate method of comparison has been devised by Rutherford and Chadwick.[1] The method consists in balancing the current produced in an ionisation vessel by the γ-rays from the two specimens of radium to be compared in turn against that produced in a second vessel by uranium placed inside it. The ionisation in this vessel remains constant whilst that in the other one is varied by altering the distance of the radium from it. When the ionisation currents have been balanced the position of the radium on a scale gives a measure of the quantity present.

The disposition of the apparatus is shown diagrammatically in *Fig.* 2. In the apparatus described by Rutherford and Chadwick, the vessel A consists of a lead cylinder 2 cm. thick, 10 cm. long, and of internal diameter 15 cm. The end of the vessel, through which the γ-rays enter, is covered with a lead plate 1 cm. thick. This vessel is placed at one end of a graduated scale about four metres long along which a slide carrying the radium R can be moved to any desired distance from it. The insulated electrode of the chamber is connected to that of the chamber B by a wire which is enclosed in an earthed tube T, about

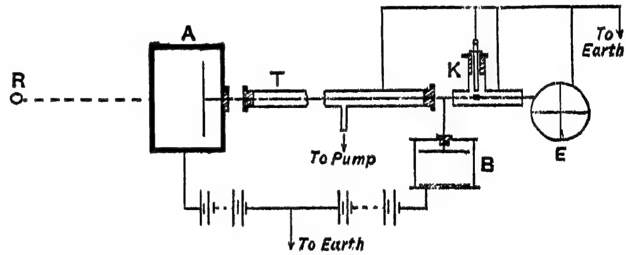


FIG. 2.

3 to 4 metres long, which can be evacuated. This prevents the production of ionisation in the tube by stray γ-rays which would vitiate the results. The vessel B is a brass cylinder 4 cm. high and 4 cm. in diameter. The ionisation in this vessel is produced by the α-rays emitted by a film of uranium oxide deposited on a plate placed at the bottom of the chamber. The ionisation can be varied over a large range by opening or shutting an iris diaphragm placed above the uranium. The insulated system is connected through

[1] *Phys. Soc. Proc.*, 1912, xxiv. 141.

an earthing key K to one pair of quadrants of the electrometer E, the other pair of quadrants being earthed. In order to produce ionisation currents in opposite directions in the two chambers, the one chamber is connected to a potential (usually about 200 volts) equal and opposite to that on the other. The electrometer and ionisation chamber B are both enclosed in a box covered with lead 5 mm. thick, and as a further protection from the γ-rays, the vessel B is enclosed in a lead cylinder 6 cm. high and 3 cm. thick, and insulated from it. It is convenient to arrange the apparatus so that the movement of the spot of light on the scale can be watched at the same time that the earthing key is worked and the slide carrying the radium is moved along the scale.

In making a comparison of two preparations of radium, the only measurements required are the distances of the balancing points from the surface of the chamber. It is necessary, however, to know the depth of the ionisation chamber A, and a correction has to be made for the absorption of the γ-rays in the air between the sources and the vessel A. In the calculation use is made of the fact that over a wide range of distance the intensity of the radiation producing ionisation is inversely proportional to the square of the distance of the radium from the ionisation vessel.

If r is the distance of the radium from the front surface of the ionisation vessel and t the thickness of the lead plate covering it, then if μ_1 is the absorption coefficient of the γ-rays in lead and μ their absorption coefficient in air, the intensity of the radiation in the ionisation vessel is proportional to $e^{-\mu_1 t}\cdot e^{-\mu r}/r(r+d)$, where d is the depth of the ionisation vessel, which is small compared with r.

Hence if R_1 milligrammes of radium, placed at a distance r_1, produce the same ionisation in the vessel as R_2 milligrammes placed at a distance r_2, then

$$\frac{R_1 e^{-\mu r_1}}{r_1(r_1+d)}=\frac{R_2 e^{-\mu r_2}}{r_2(r_2+d)},$$

i.e.

$$\frac{R_2}{R_1}=\frac{r_2}{r_1}\cdot\frac{r_2+d}{r_1+d}\cdot e^{-\mu(r_1-r_2)}.$$

This method can be employed to measure quantities of radium greater than one milligramme to an accuracy of about 1 part in 400. Smaller quantities than this have to be brought so near to the ionisation vessel that the inverse square law no longer holds.

(*c*) *Mme Curie's Method.*—A γ-ray method due to Mme Curie is shown in *Fig.* 3. The radium is placed on top of a large circular plate condenser C consisting of two sheets of lead about 80 cm. diameter and 5 mm. thick. An insulated aluminium plate situated between the lead plates, with a clearance of about 2 millimetres on either side, serves as the electrode. To obtain saturation the lead plates are connected to a high potential. The ionisation current produced is balanced by the use of a quartz piezo-electrique Q. The relative saturation currents produced by the test and standard preparations afford a definite measure of the radium present.

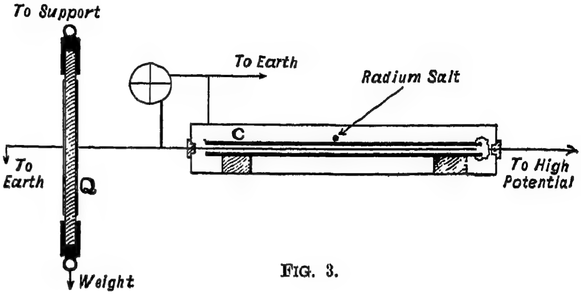


FIG. 3.

(*d*) *Soddy's Method.*—Soddy describes a γ-ray method by which the radium content of a low-grade preparation can be estimated.[1] A spherical flask is filled with a known weight of the salt and sealed up. After equilibrium has been attained a comparison is made between this and the standard by the first method described above. An accurate correction can be made in this instance for the absorption of the rays in the salt. Suppose I is the intensity of the radiation observed from the flask and I_0 what it would be if no absorption of the radiation occurred. Then it can be shown that

$$\frac{I}{I_0}=1-\tfrac{3}{4}\mu R+\tfrac{4}{10}(\mu R)^2-\tfrac{1}{6}(\mu R)^3+\tfrac{2}{35}(\mu R)^4-\tfrac{1}{60}(\mu R)^5+\ .\ .\ .\ ,$$

where μ is the absorption coefficient of the radiation in the material and R the radius of the flask. The value of μ/ρ where ρ is the density may be taken as 0·040, the density of the substance being taken as the mass divided by the known volume V of the flask. The series is rapidly converging, and three or four terms will suffice in most cases.

For a rapid approximate estimation of radium in minerals Soddy employs a large lead electroscope, about 20 cm. high by 20 cm. diameter and of wall thickness 0·4 cm.

[1] Soddy, *The Chemistry of the Radio-elements*, 1917, i. 93.

The windows are made of lead glass and are shielded by lead so as to prevent the entry of secondary radiation into the electroscope. The leaf system is of the simple type and is insulated by sulphur. In comparing pitchblendes and other minerals a lump of 20 to 100 grammes, according to the richness of the mineral, is taken and placed on top of the electroscope. The increase in the leak is compared with that produced by a similar lump of a standard pitchblende, the uranium content of which is known. These measurements supply enough data to calculate the amount of radium in the mineral. For example:

	Per cent. U	Weight.	Leak of Electroscope.
Standard . .	60	21 gms.	12 div. per min.
Specimen tested .	x	30 ,,	8 ,, ,, ,,

$$x = \frac{21}{30} \times \frac{8}{12} \times 60 = 28 \text{ per cent.}$$

The "equilibrium ratio" of radium to uranium in a mineral of old formation is $0{\cdot}323 \times 10^{-6}$, that is, there are 323 milligrammes of radium element present per ton of uranium element. Hence the amount of radium in the above sample is $0{\cdot}28 \times 0{\cdot}323$, *i.e.* 0·090 gramme per ton of mineral.

(ii.) *The Emanation Method.*—This method is very suitable for accurate measurements of minute quantities of radium. It depends on the fact that radium produces a gaseous product, the emanation, of a comparatively long period of transformation which can be separated completely from radium solutions. The emanation reaches its equilibrium value after the radium solution has been sealed up in an enclosure for about a month; when this stage is reached the amount of emanation present is proportional to the radium content of the solution. As a standard of comparison it is necessary to prepare solutions containing a known quantity of radium which has been measured by one of the γ-ray methods. A standard solution is prepared as follows: A known quantity of radium salt, say 1 milligramme, is dissolved in distilled water containing a little dilute hydrochloric acid so as to get it all into solution. This is made up to a litre by the addition of water. One cubic centimetre of this solution is measured carefully, either by means of a pipette or by weighing, and transferred to another flask, and water added up to a known volume, say 1 litre: 1 cubic centimetre of this solution would therefore contain 10^{-6} milligramme of radium. A number of solutions of different strengths are prepared in this manner and kept as standards of reference. It is essential to take great precautions that all the radium is initially put into solution, and it is advisable to add a little hydrochloric acid to the standard solutions when they are made, so as to ensure that the radium remains permanently dissolved. After the standard solution is made, it is placed in a distilling flask which can be sealed up. Before sealing up, however, the solution is thoroughly boiled to drive off all the emanation. Boiling for a few minutes suffices for this purpose. The flask is sealed up before it cools and is kept for a month to allow the emanation to accumulate.

When the radium content of a sample is required, it is put into solution in a manner similar to that described above in the case of the standard. In a month after the vessel containing the solution is sealed, the emanation is transferred into an electroscope. This transference is effected by means of the apparatus shown in *Fig.* 4. The distilling

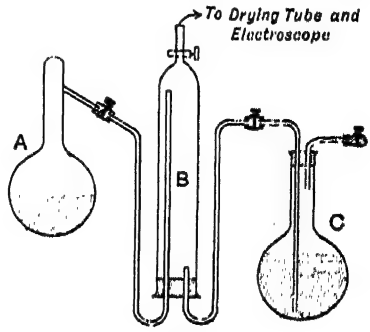


Fig. 4.

flask A, containing the radium solution, is opened and the air rushes into the partially exhausted vessel so that there is no loss of emanation. This is then connected to the gas burette B, which is filled with hot water forced into it from the flask C. The flask A is then heated for about five minutes so as to drive the emanation into the gas burette; the emanation collects in the top of the burette and does not dissolve in the hot water. Cold water dissolves the emanation to an appreciable extent. The flask A is again sealed, and the emanation in the burette sucked into an electroscope where its activity is measured. To do this the electroscope is evacuated by means of a water pump and the emanation passed into it through a drying tube. The burette is finally disconnected from the drying tube and a stream of dry air allowed to pass into the electroscope until atmospheric pressure is established. The electroscope is then closed. The rate of leak due to the emanation in the electroscope increases rapidly at first and then more slowly, reaching a maximum in about three hours

after the introduction of the emanation. The increase of leak is due to the production by the emanation of its products — radium A, radium B, and radium C. During this interval it is advisable to keep the leaf of the electroscope negatively charged so as to concentrate the active deposit on the central rod. Owing to the intense field near the central electrode, saturation is obtained more easily than if the active deposit were driven to the walls of the electroscope. For accurate work it is desirable to take measurements of the leak when at its maximum. The emanation should be swept out of the electroscope as soon as the measurements are made, so as to avoid contamination from radium F, which will increase the natural leak.

The same procedure is gone through with the standard solution. The ratio of the leaks in the two cases gives a measure of the amount of radium in the sample in terms of the standard.

The emanation method is very accurate and reliable. The emanation from 10^{-6} milligramme of radium gives a comparatively rapid discharge, and it is possible to measure one-hundredth of this quantity with certainty.

§ (10) Mesothorium. — The γ-ray method does not differentiate between radium and mesothorium, both of which emit penetrating γ-radiation. The penetrating power of the rays from radium is slightly greater than that of the rays from mesothorium, and a careful comparison of the absorption of the rays in lead will reveal the contamination of radium by mesothorium without the necessity of opening the tube containing the preparation. A more satisfactory, although a more laborious method of detecting the presence of mesothorium is to dissolve the salt and drive off the emanation. If the solution contains no mesothorium the γ-ray intensity will almost vanish after three hours; this will not be the case if there is any mesothorium present. The objection to this method is that the tube containing the preparation must be opened, which involves the risk of the loss of some of the material.

The carnotites are known to contain only a negligible amount of mesothorium. Tests made at the Bureau of Standards, Washington, on radium produced from this ore indicated that the mesothorium present did not exceed 0·2 per cent of the radium content of the material. Consequently it is quite safe to assume that the radium produced from these deposits is practically free from mesothorium.

§ (11) Radium Constants.—Radium itself emits $3{\cdot}57 \times 10^{10}$ α-particles per second per gramme of radium; when in equilibrium with its products $14{\cdot}3 \times 10^{10}$ are expelled per second per gramme of radium. The total charge carried by the α-particles emitted per second from 1 gramme of radium and from each of its products in equilibrium with it is 33·2 e.s.u. or $1{\cdot}11 \times 10^{-9}$ e.m.u.[1]

The quantity of emanation in equilibrium with 1 gramme of radium is known as a *curie* of emanation, and one-thousandth part of it is a *millicurie*. Similarly the quantity of radium A, radium B, and radium C in equilibrium with one gramme of radium are known as one curie of radium A, radium B, or radium C respectively. The total ionisation current due to α-rays from one curie of emanation is $2{\cdot}89 \times 10^6$ e.s.u. when the emanation is by itself and $9{\cdot}94 \times 10^6$ e.s.u. when it is with its α-ray products.[2] The total charge carried by the β-particles emitted per second by Ra B or Ra C in equilibrium with 1 gramme of radium is 18·3 e.s.u.[3] The volume of the emanation from 1 gramme of radium in equilibrium is 0·63 cubic millimetre, which agrees closely with the value calculated from the knowledge of the number of α-particles emitted per second, per gramme of radium itself, namely, 0·62 cubic millimetre.[4] The total heating effect of 1 gramme of radium and its products in equilibrium with it is 134·7 gramme calories per hour. The contribution to this total amount by each of the products is as follows:

Radium alone .	. 25·1	gramme calories per hour.
Radium emanation .	28·6	,, ,, ,,
Radium A . .	. 30·5	,, ,, ,,
Radium B } Radium C } .	. 50·5	,, ,, ,,

The production of helium by radium was investigated by Ramsay and Soddy.[5] Purified emanation from about 50 milligrammes of radium bromide was introduced into a small spectrum tube. No helium could be detected spectroscopically immediately after the introduction of the emanation, but after standing for four days the helium spectrum appeared with all its characteristic lines. This showed that helium is produced directly by the transformation of the radium emanation. It was the first definite evidence of the production of a known element during the transformation of radioactive matter. Boltwood and Rutherford[6] found experimentally that 164 cubic millimetres of helium are produced per gramme of radium per year; this figure is in good agreement with that found by calculation, namely, 163 cubic millimetres.

From the number of α-particles emitted per second by 1 gramme of radium Rutherford and Geiger calculated that the half-value

[1] Rutherford and Geiger, *Roy. Soc. Proc.*, 1908, A, lxxxi. 141.
[2] Geiger, *ibid.*, 1909, A, lxxxii. 486.
[3] Moseley, *ibid.*, 1912, A, lxxxvii. 230.
[4] Rutherford, *Phil. Mag.*, 1908, xvi. 300.
[5] *Roy. Soc. Proc.*, 1903, A, lxxii. 204.
[6] *Phil. Mag.*, 1911, xxii. 586.

period of transformation of radium was 1690 years. Boltwood[1] determined this quantity experimentally by separating the whole of the ionium present in a uranium mineral and determining the growth of the radium from it in terms of the equilibrium amount of radium in the mineral. He found that the half-value period was about 2000 years. A more recent determination by this method, and employing four different minerals, has been made by Mlle. Gleditsch.[2] The mean of the four determinations gave the value 1660 years for the half-value period, which agrees more closely with the above-calculated value. If I_0 is the initial activity of radium and I its activity after an interval of time t, then $I/I_0=e^{-\lambda t}$, where λ is a constant known as the "radioactive constant." It may be defined as the fraction of the total amount disintegrating per unit of time, assuming the time unit to be so small that the quantity at the end of the time unit is not sensibly different from that at the beginning. For radium $I/I_0=0\cdot5$ when $t=1690$ years, so that $\lambda=4\cdot10\times10^{-4}$ per year. An atom disintegrating according to the above law can be shown to have an average life of $1/\lambda$, which in the case of radium would be 2440 years.

BIBLIOGRAPHY

E. Rutherford, *Radioactive Substances and their Radiations*, Cambridge University Press, 1913; W. Makower and H. Geiger, *Practical Measurements in Radioactivity*, Longmans, Green & Co., 1912; F. Soddy, *The Chemistry of the Radio-elements*, Part I., Longmans, Green & Co., 1917.

E. A. O.

Radium, Principal Commercial Ores of, pitchblende and carnotite. See "Radium," § (2).

Radium Active Deposit, Decay of. See "Radioactivity," § (20) (ii.).

Radium Content of Luminous Compound, measurement of, by γ-ray method. See "Luminous Compounds," § (5).

Radium Emanation: a gaseous product of the disintegration of radium; employed as the basis of an accurate method of measurement of minute quantities of radium. See "Radium," § (9) (v.).

Radium Luminous Compound, Photometry of. See "Photometry and Illumination," § (125).

Radium Painted Dials, Photometry of. See "Photometry and Illumination," § (127).

Ramsden's Eyepiece. See "Eyepieces," § (3); "Telescope," § (6).

[1] *Am. J. Sci.*, 1908, xxv. 493.
[2] *Ibid.*, 1916, xli. 112.

RANGE-FINDER, SHORT-BASE

§ (1) **Early History of the Range-finder.**—The evolution of the short home-base range-finder provides one of the most interesting stories in the whole range of applied optics. From the earliest times, the distances of inaccessible objects have been determined by simple triangulation methods. This has resulted in a constant demand for instruments of the theodolite type, designed to measure angles in azimuth with the greatest possible accuracy. When, however, the problem of determining ranges during a battle, under the eyes and fire of the enemy it might be, became an urgent one, it was soon discovered that what were in essence surveying methods suffered under great disadvantages. They necessitated the use of a comparatively long base, and the employment of a number of men, for the carrying out of an operation which took a considerable time, and had to be conducted under conditions which made accurate work difficult. Thus arose the demand for the short-base, single-observation range-finder.

In the earlier forms of short-base range-finders two theodolites in effect were mounted at the ends of a known and rigid base. The theodolite telescopes were directed either successively by a single observer, or simultaneously by different observers, on to the distant object, and the convergence angle determined from the readings of the theodolites.

In later forms of this type the two telescopes were fixed rigidly to the ends of the base with their axes parallel to one another and at right angles to the base. In making an observation the image of the object was brought on to the fixed cross-wires of one telescope, by rotating the range-finder in azimuth, and then the displacement of the image from the axis in the second telescope was determined by an ocular micrometer, graduated, it might be, to give ranges. This marked a decided advance. In effect the object was brought into position along the axis of one telescope, and the range determined from the parallactic displacement of its image in the field of view of the other telescope.

The two-independent-telescope type, however, was open to manifest objections. When two observers were employed it was difficult to ensure that the same object was being ranged upon in each telescope, and ranging upon a moving object presented even greater difficulties. It was thus realised that the two telescopes employed should project their images into a single common eyepiece, so that the two images given by the two telescopes could be observed simultaneously by a single observer.

One of the earliest forms of coincidence, or

rather superposition range-finders, was that invented in 1775 by an optician named Magellan. In this apparatus, as shown by *Fig.* 1, sliding telescopic tubes *a* and *b* carry mirrors M′ and M at their outer ends, the latter mirror being fixed in front of a telescope L. Thus a distant object can be seen directly in the telescope L, and indirectly, after suc-

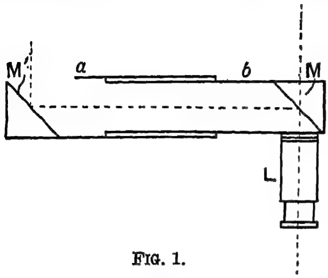


FIG. 1.

cessive reflection of the light by mirrors M′ and M. These mirrors are inclined at a small angle to one another so that within certain limits superposition of the image can be effected by adjustment of the base length. From this base length and the constant angle of convergence of the mirrors the range can be determined. A notable instrument of this type was also invented for military purposes by Brander, a description of which was published in 1781.

An interesting range-finder was that patented in 1858 by General Clark (see *Fig.* 2). In this instrument a single telescope only was employed. Light from the object fell simultane-

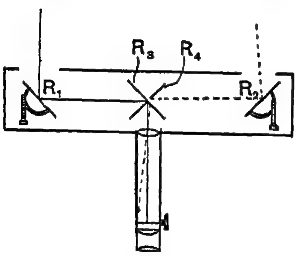


FIG. 2.

ously upon two end reflectors R_1 and R_2, from which the light passed to inclined reflectors R_3 and R_4 mounted in front of a telescope object-glass. Two pictures, displaced horizontally with respect to one another, were thus seen in the focal plane of the eyepiece. This horizontal displacement, which was a measure of the range, was measured directly by means of a micrometer screw.

It should be noted that Clark's instrument was not, strictly speaking, a coincidence range-finder, since the actual displacement of one picture with respect to the other was determined directly. It was, however, one of the earliest instruments of fixed base length which permitted of simultaneous observation by a single observer in a single eyepiece.

In 1860 Adie took out what would now be called a master patent for a short, constant base, coincidence range-finder. Up to this time coincidence, or superposition, had been secured by varying the base length ; or, the base length being fixed, the parallactic displacement in the ocular field had been measured directly. *Fig.* 3 is taken from Adie's specification. An outer tube *d* carries a telescope system consisting of an object-glass *b′*, prisms p^2 and p^3, and an eyepiece a^2. An inner tube *d′* pivoted at *c* carries a second telescope object-glass *a′* and prisms *p*, *p′*. Each of the telescopes therefore projects an image into the focal plane of the eyepiece a^2. When the inner tube is rotated about its pivot *c* by the operation of a micrometer d^2, the image produced by the telescopic system on the left is moved horizontally across the

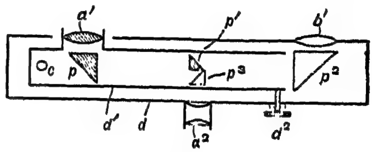


FIG. 3.

line of sight in the field of view, and can thus be superposed, or brought into coincidence with the image produced by the telescope on the right, and the range determined. With such a simple system of reflecting separating prisms, partial fields with a sharp dividing line cannot be obtained, a condition which has been shown to be essential to accurate work.

An interesting specimen of Adie's telemeter is to be seen in the Science Museum, South Kensington. It has a base length of 3 ft. The parallactic angles are read off directly, whilst reference to a table carried by the instrument gives the corresponding ranges from 100 to 2000 yards. The instrument was used on H.M.S. *Triton*, from 1885 to 1904, during the survey of the East Coast of England.

Very soon after the date of Adie's invention, the German optician Steinheil, and the French optician Tavernier, brought out coincidence range-finders in which the coincidence was brought about by a rotation of one of the end reflectors. Marindin as late as 1901 patented a range-finder in which coincidence was secured by a similar tilting of an end reflector.

In 1868 Paschwitz described a range-finder in which coincidence was obtained by swinging a plane-parallel plate in the beam passing along

one of the telescopes, that is in a convergent beam.

The pentagonal prism appears to have been first applied to range-finders by Colonel Goulier of the French Army about 1864.

In 1885 Mallock introduced a range-finder in which an optical square made up of two plane reflectors, inclined at 45° to one another, was used as an end reflector. Coincidence was brought about by tilting a reflector by means of a micrometer.

Finally, in 1888, Messrs. Barr & Stroud invented their well-known range-finder, which embodied in one instrument so many of the features which experience has shown to be essential for efficiency in a service instrument.

Thus far we have dealt with the history of the coincidence or monocular type of range-finder only. Quite early, however, as we shall see, it was recognised that a second type, depending upon the plastic character of binocular vision, was possible.

§ (2) STEREOSCOPIC RANGE-FINDERS.—In binocular free vision, objects can be relatively positioned in the line of sight with very considerable accuracy, but a limit is imposed by the fact that before an object can be stereoscopically resolved, or seen in relief, it must have a certain minimum dimension in the line of sight, a dimension which, according to Helmholtz, must subtend angles at the eyes of the observer, the sum of which is at least equal to one minute of arc. This limiting angle, it may at once be said, has been reduced to 20 seconds of arc, and even less, by later workers. The sum of these angles, it can easily be shown, measures the change in the convergence angle of the axes of the eye when directed first at the nearer, then at the more remote end of the object; thus, put in other words, the statement means that in order that two objects at different distances from an observer shall be seen as at different distances, the convergence angle of the axes of the eyes must change by at least 20″ in passing from one object to the other.

Now assuming an average inter-pupillary distance of $2\frac{1}{2}$ inches, the distance at which this inter-pupillary distance subtends an angle of 20″ is equal to $2\frac{1}{2}/\tan 20''$ inches, a distance of about 700 yards, whilst for an object at infinity the convergence angle is of course zero. Obviously, then, no two objects which occur in space outside a sphere with a radius of 700 yards can be stereoscopically resolved, *i.e.* the radius of the spherical stereoscopic field of free vision is about 700 yards. Nothing beyond that can be seen in relief as compared with an object even at an infinite distance. A point of light, for instance, beyond 700 yards away cannot be seen to be in relief against stars in the sky.

Accepting the principle of Helmholtz's limit, it is obvious that the greater the inter-pupillary distance of an observer, the greater must be the radius of his stereoscopic field of view. Herschel and Helmholtz, independently of one another, very ingeniously took advantage of this fact about the years 1855–58, and by optically magnifying the inter-pupillary distance of an observer, gave to him increased stereoscopic power. This was done by a reflecting system of four mirrors R_1, R_2, R_3, R_4 (*Fig.* 4). Light from a distant object point

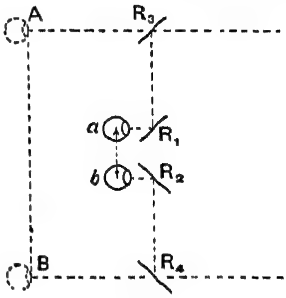


FIG. 4.

reaches the eyes respectively by successive reflections from the mirrors R_3 and R_1, and R_4 and R_2. The effect of this is that the inter-pupillary distance *ab* is optically magnified to the distance AB, with a corresponding increase in stereoscopic power. The limit of 20″ now becomes $(20 \times AB)/ab$ seconds. Herschel and Helmholtz both proposed, further, to combine the reflecting system shown with a binocular telescope, and thus obtain a further increase in the final convergence angle, and thus, it may be remarked, a further increase of the radius of the spherical stereoscopic field of view.

A remarkable reference to a stereoscopic range-finder occurs as early as 1738, in Smith's *A Compleat System of Optiks*.[1] The instrument and its use is described in the following words:

"Having opened the points of a pair of compasses, somewhat wider than the interval of the eyes, with your arm extended, hold the head, or joint, in the ball of your hand, with the points outwards, and equidistant from your eyes, and somewhat higher than the joint. Then fixing your eyes upon any remote object lying in the line that bisects the interval of the points, you will first perceive two pairs of compasses (each leg being doubled) with their inner legs crossing each other, not unlike the old shape of the letter W. [The old W referred to was like two V's slightly overlapping.] But by compressing the legs with your hands, the two inner points will come

[1] See *A Compleat System of Optiks*, vol. ii. p. 388.

nearer to each other, and when they unite (having stopped the compression) the two inner legs will also entirely coincide and bisect the angle under the outward ones, and will appear more vivid, thicker and longer than they do, so as to reach from your hand to the remotest object in view, even in the horizon itself, if the points be exactly coincident."

We have in the words quoted a very complete and interesting anticipation of what we shall subsequently refer to as the "wandering-mark" type of stereoscopic range-finder. It is, however, to Hector Alexander de Grousilliers, who took out a patent in 1893 for an "Improved Stereoscopic Telemeter," that the modern instrument is due. The specification in question is a remarkable document. It sets out the theory of the instrument in the following words:

"Assuming A_1A_2 to be points at which, in both fields of view, an infinitely distant object A is depicted, and N_1N_2 to be the points at which is depicted an object N, situated at a finite distance E from the observer; the difference between the linear distances A_2, N_2 and A_1, N_1, *i.e.* the parallactic difference between the image of the one object, situated at an infinite distance, and the other object, placed at the finite distance E, will be

$$A_2N_2 - A_1N_1 = \delta = \frac{\Delta}{E} \cdot f,$$

f being the equivalent focal length of the telescope-lenses, and Δ the distance between their respective axes.

"In the case of two objects N and N′, both of which are located at finite distances E, E′ from the observer, the parallactic difference in the two images will consequently be

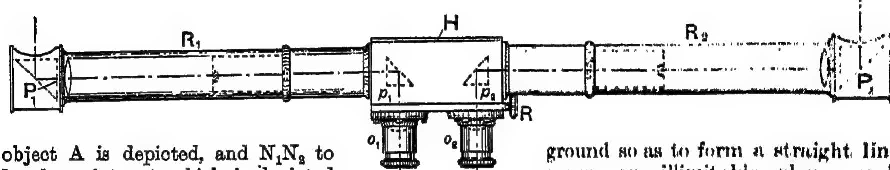
$$N_2N_2' - N_1N_1' = \left(\frac{1}{E} - \frac{1}{E'}\right)\Delta \cdot f."$$

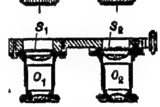
Finally Grousilliers describes and claims three types of the range-finder, namely: (1) the *fixed scale*, in which each eyepiece has a fixed scale, which bears for each range a range mark corresponding to a similar mark on a similar scale in the other eyepiece; (2) the *wandering-mark*, in which a single fixed mark occurs in one eyepiece, and a similar mark, but capable of micrometric displacement in the plane of the instrument and across the line of sight, in the other; and (3) what may be called the *wandering-picture* type, in which one of the images, or pictures, of the object may be displaced in a similar way to the wandering-mark in the second type, to bring the stereoscopic image of the object into the same plane as that of the stereoscopic image of two fixed marks, one in each eyepiece.

In effect, in the first type, when a range is taken, the pair of corresponding scale marks is selected which has the same inter-eyepiece separation as the two images of the object point under observation. In the second type, the wandering-mark is adjusted to obtain this equality of separation; whilst in the third type, one of the pictures is traversed across the line of sight until the separation of the two images of the object point is made equal to the fixed separation of the two range marks.

The fixed scales required in the first type of range-finder might be produced in the following way:

Let a number of pegs be driven into the ground so as to form a straight line across an illimitable plane, and, further, let each peg be marked with its distance from a certain place. Now suppose that from this place a pair of stereoscopic photographs of these pegs are taken with a binocular prismatic telescope, from which the oculars have been removed, to allow of photographic plates being fixed in the two focal planes of the objectives. Those photographs, after development, could be returned to the binocular telescope so as to occupy the positions that they did during the taking of the photograph. In the field of a binocular, equipped with scales produced in this way, and directed to a distant landscape, that landscape would be seen in stereoscopic relief in the usual way, and passing through it, or over it, would be seen, also in relief, a distance scale, to a position along which any object might be assigned and its distance thus determined. Stereoscopic scales are not, as a matter of fact, usually prepared in this way. It is more convenient to calculate and rule them.





FIGS. 5 and 6.

Figs. 5 and 6 are taken from Grousilliers' specification. Two broken telescopic systems are mounted in tubes R_1 and R_2, secured collinearly to the ends of a box-like casing H. End reflectors P_1 and P_2 deflect the light along the axes of these telescopes, whilst reflectors p_1, p_2 reflect the beams outwards parallel to one another, through eyepieces o_1 and o_2, in the focal planes of which stereoscopic

scales s_1 and s_2 are mounted, the one on the right being movable by means of a micrometer R, to effect the necessary adjustment and thus determine the range. Since the telescopes are provided with ordinary erecting systems the reflections by P_1, p_1 cancel one another, with the result that the images are seen erect.

Grousilliers' claim for the fixed-scale modification is as follows:

"An improved telemeter, consisting of a double telescope (which term includes telestereoscopes with an enlarged distance between the object-glasses) within the ocular fields of which scales are so arranged that, when viewed binocularly, they present the image of a series of distance marks extending downwards and situated at predetermined distances from the observer: the image of such series of marks being projected into the image or panorama of a landscape, as seen through the double telescope, so that the distance of any object viewed may be immediately read off the scale according to the point between any two marks of such scale to which the object can be made to correspond with regard to the visual depth."

§ (3) Theory of the Short Home-base Range-finder.—The theory of the range-finder is based upon a method—well known to schoolboys—for finding the distance of an inaccessible object.

Let A, *Figs.* 7 and 8, be the object, and let a rectangle BDEC, with sides of length b and f,

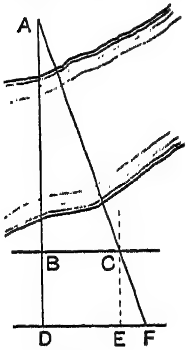


Fig. 7.

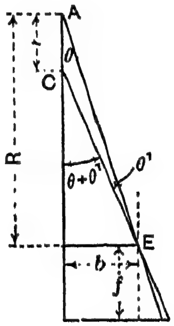


Fig. 8.

be so positioned that the object A is seen along the lines DB and FC, F being in line with C and A. With the object A at an infinite distance, it would be sighted along the two parallel lines DB and EC. The length EF, equal to d, say, is therefore the parallactic displacement along the line DE of the intersection of the line of sight CA, whilst BAC is the corresponding parallactic angle, *i.e.* the angle subtended at the object point A by the base b.

Now, since the triangles ABC and CEF are similar, it follows that R, the range, equal to BA, is to b as f is to d, that is, the distance of the object A is equal to the product bf, divided by the parallactic displacement. This is the fundamental equation for both coincidence and stereoscopic range-finders:

$$R = \frac{bf}{d}.$$

In the case of both these range-finders, b represents the optical base length of the instruments, f the focal length of each of the two telescope objectives, R the range, and d the parallactic displacement in the common focal plane of the two objectives when the range-finder scale is set to infinity. The accuracy of the range-finder then depends fundamentally upon the accuracy with which the parallactic displacement d can be measured, and this in the limit depends on the eye.

§ (4) The Accuracy of Range-finding as determined by the Eye. (i.) *Coincidence Type.*—In this type the accuracy of the results depends upon the aligning power of the eye, as exemplified, for example, in the setting and reading of a vernier. In the case of the range-finder, this aligning takes place usually across a sharp horizontal line, which divides the field of view into two parts. A flagstaff, for instance, that is being ranged upon, is seen with its top part, carrying the flag, displaced laterally with respect to the bottom. In adjusting for coincidence these parts are brought into alignment, so that the accuracy with which this operation can be done determines the accuracy of the range, other things being equal. The important question then is—given two collinear straight lines in the field of view, through what angle a can one be moved laterally with respect to the other before the fact that they are no longer collinear is detected by the eye? For practical range-finding purposes this angle has been taken as equal to about 12 seconds of arc in free vision, and the same, therefore, in the image-space, or final field of view presented by a range-finder. There is no doubt, however, that under favourable conditions many trained observers can work down to an angle of the order of 5 or 6 seconds of arc. Further, on artificial objects such as lines drawn on silvered glass plates placed in the focal planes of collimators, higher accuracy even than this can be obtained — 2 to 3 seconds of arc. Now let θ' (*Fig.* 8) represent in actual space—the object-space—the angle corresponding to a in the image-space, then an object accurately ranged upon in the position A may move up through a distance r, subtending the angle θ', at the range-finder,[1] until the angular parallactic

[1] For the purpose of simplifying the diagrams we have supposed the object to move up along the axis of one of the telescopes.

displacement in the field of view is equal to α, before an object at C can be differentiated, with respect to its distance, from an object at A. The possible error under these circumstances in ranging upon an object at A is equal to the distance r, and it may obviously occur in either direction.

(ii.) *Stereoscopic Type.*—In this case the following considerations apply: If two pins be stuck into a table close together, and at the same distance from an observer, then one pin can be moved backwards and forwards in the line of sight through a certain distance before the observer can see with certainty that they are no longer at the same distance away; that is, a certain change in the angle of convergence of the optic axes of the eyes is necessary in order that two objects can be seen to be at different distances. Helmholtz, as mentioned above, made a number of experiments in this way and came to the conclusion that in free vision a difference between the convergence angles of one minute of arc was the least difference necessary for differentiation of distance in the line of sight. Later workers have, however, reduced this limit to something of the same order as that claimed for the coincidence range-finder. We shall revert to the question later.

(iii.) *Coincidence and Stereoscopic Types.*—We will assume, therefore, that the possible error angle α represents (i.) in the case of the coincidence range finder, the greatest angular parallactic displacement in the field of view which can occur without loss of coincidence being detected; whereas (ii.) in the case of the stereoscopic range-finder, it represents the maximum difference of the convergence angles in the field of view which can occur without apparent change of the range of the object under observation.

§ (5) FUNDAMENTAL RANGE-FINDING EQUATIONS.—In *Fig.* 8 the range-finder with base length b subtends at the object A the angle θ. Whilst the range changes by a length r, the angle of parallax changes by an angle θ'. To find the relationship between these changes, the angles θ and θ' being expressed in circular measure. Now

$$\theta=\frac{b}{R}, \quad . \quad . \quad . \quad . \quad . \quad (1)$$

and since in any triangle the sides are proportional to the sines of the opposite angles, and in the triangle ACE the angles θ and θ' are small, we have

$$\frac{CA}{CE}=\frac{\theta'}{\theta},$$

and since CA $=r$ and CE (very nearly) $=$R,

$$\frac{r}{R}=\frac{\theta'}{\theta}. \quad . \quad . \quad . \quad . \quad . \quad (2)$$

Substituting the value of θ from (1),

$$r=\frac{R^2\theta'}{b}. \quad . \quad . \quad . \quad . \quad . \quad (3)$$

Let P_e be the percentage error for a given range R, then

$$P_e=\frac{100r}{R},$$

and substituting the value of r from (3),

$$P_e=\frac{100R\theta'}{b}. \quad . \quad . \quad . \quad . \quad (4)$$

Let α be the error angle, as set out above, for both the coincidence and the stereoscopic types of range-finders, and M the magnifying power of the telescopic systems employed; then, in the case of the coincidence range-finder (see *Fig.* 8), it is evident that the length AC, in the object-space, subtending an angle θ', at the window of the range-finder, will appear to the observer, *i.e.* in the image-space, under an angle $M\theta'$; whilst in the case of the stereoscopic range-finder, the angles of convergence in the object-space, which are θ for the point A, and $\theta+\theta'$ for the point C, become $M\theta$ and $M(\theta+\theta')$ respectively, in the image-space, so that the change in the convergence angles is equal to $M\theta'$. Thus for both types,

$$\theta'=\frac{\alpha}{M}.$$

Substituting in (3) and (4), we obtain

$$r=\frac{R^2\alpha}{bM}, \quad . \quad . \quad . \quad . \quad . \quad (5)$$

and

$$P_e=\frac{100R\alpha}{bM}. \quad . \quad . \quad . \quad . \quad . \quad (6)$$

Thus, α, b, and M being constants, the percentage error varies directly as the range.

To express θ' in (3) in seconds of arc, instead of in circular measure,

$$\theta^1=\frac{br}{R^2}206000. \quad . \quad . \quad . \quad . \quad (7)$$

Assuming a value for α of 12 seconds of arc, as is usually done, equation (6) enables us to formulate a very useful rule, namely, that

"The percentage error of any range-finder at one thousand yards is equal to 6 divided by the bM value," since

$$P_e=\frac{100\times 1000\times\cdot 00006}{bM}$$

$$=\frac{6}{bM}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

b being expressed in yards.

The corresponding error in range r is obtained from

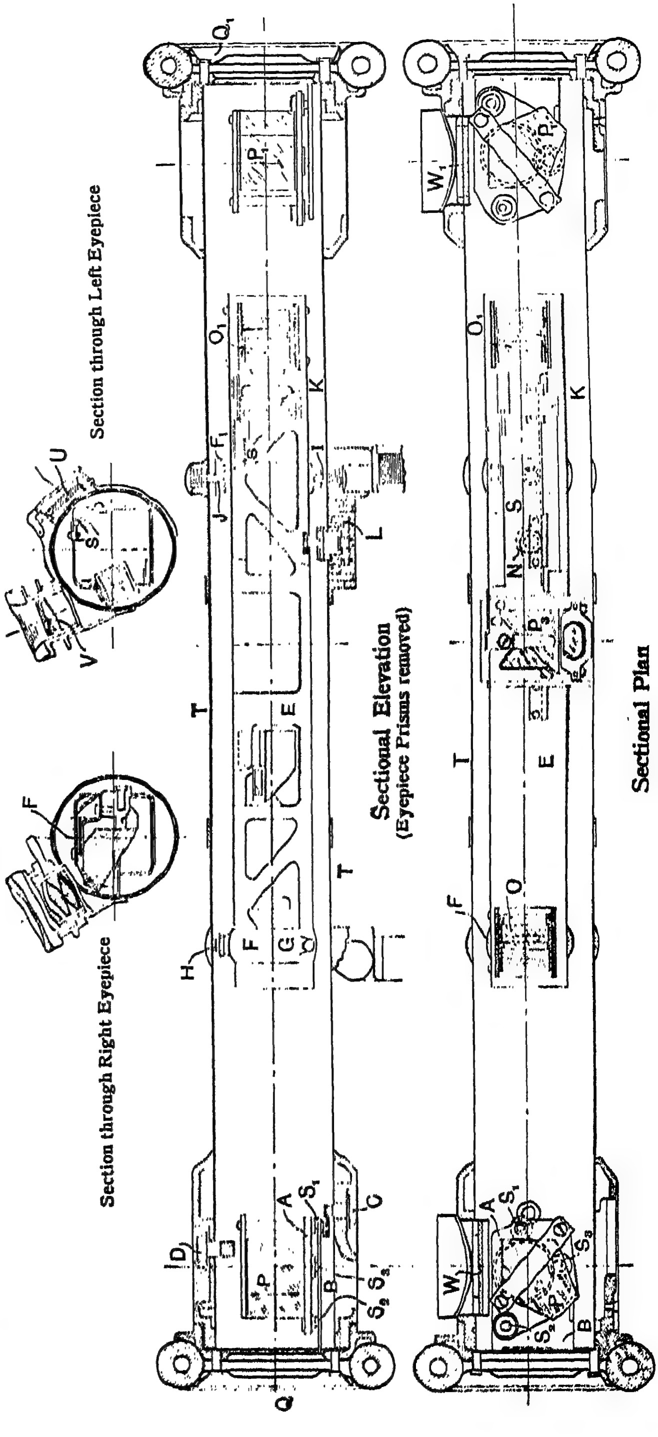
$$r=\frac{6\times 10}{bM}. \quad . \quad . \quad . \quad . \quad . \quad (9)$$

Knowing the range error r for 1000 yards, it follows from equation (5) that the range error for any other range R_t, expressed in thousands of yards, is equal to rR_t^2. Thus the error at 1000 yards being 5 yards, that at 10,000 yards would be 500 yards.

A word of caution is, perhaps, necessary here. The numeral in equation (8) assumes a value of 12 sec. for the angle α; but, in the opinion of many experts, this angle is in practice considerably greater for the stereoscopic than for the coincidence range-finder, and equation (6) shows that the percentage error for a given range-finder and a given range varies directly as the angle α.

I. Details of Construction

§ (6) General Construction of Coincidence Range-finders. (i.) *Barr & Stroud's Range-finder.*—*Figs.* 9 and 10 show in sectional elevation and plan respectively a small Barr & Stroud coincidence range-finder with a base length of 50 cm., a magnification of 7½ times, and scale for ranges from 350 to 1500 yards. The optical system, and the principal features of the mechanical design shown, are practically identical with those adopted in instruments of much greater power. The instrument consists of an outer cylindrical tube T, with end fittings carrying the windows W and W_1, and end caps Q, Q_1 carrying the supporting plates for the prisms P and P_1. The inner tube E, of rectangular cross-section, carries the objectives O, O_1 and the separating prism P_3, which deflects the light entering by



Section through Left Eyepiece

Section through Right Eyepiece

Sectional Elevation (Eyepiece Prisms removed)

Sectional Plan

Figs. 9 and 10.

the objectives O and O_1 to a common focal plane, in which the coincidence is secured by observation through the single eyepiece. The deflecting prism K slides backwards and forwards in the tube E between the O.G. O_1 and the separating prism P_3, to give a variable lateral displacement to the corresponding image, and thus effect coincidence under the action of a finger button L. The range-scale S is attached to the mount of the prism K, and slides in guides carried by the inner tube. The scale is observed through an eyepiece V and a system of reflecting prisms as shown. U is a condenser by means of which the scale is illuminated from the back.

It is important that any strain of the outer tube T should not be communicated to the inner tube E. To secure this end the tube E is fitted with rings F and F_1, the former of which carries, at approximately 60° apart, two fixed radial pins G and a spring pin H, engaging with the inner surface of the outer tube. By this arrangement the tube is left free to move longitudinally whilst secured co-axially with the outer tube. The second ring F_1 carries on its lower side a cup, which engages with a stud I, whilst the upper part of the rim carries a longitudinal slot, which engages with the stud J. The stud I prevents longitudinal displacement of the tube E, whilst the stud J prevents any lateral displacement of the axis of the tube E.

The adjustment of the range-scale for an object at infinity is obtained by rotating in its own plane the window W, which is in the form of a weak refracting prism with its refracting edge horizontal, by means of a finger pinion D. The prism P is clamped to a plate A, secured to the base-plate B, carried by the end-cap Q, by screws S_2 and S_3, which permit of a rocking motion of the plate A and prism P with respect to the plate B, about an axis passing through the points S_2 and S_3. The forward part of this plate A can be raised and lowered by means of a screw S_1 operated by means of a finger pinion C, to raise and lower the image in the corresponding partial field of view and thus effect the "halving" adjustment. This adjustment secures, when coincidence has been effected, that the image on one side of the separating line is a complete reflected image of that on the other, or, in the case of a circular object, like the moon, that the image is single and circular.

(ii.) *Cooke & Sons' Range-finder.* — Mr. Dennis Taylor, of Messrs. Cooke & Sons, has done a large amount of work in the development of a type of range-finder which differs from that already described in several important particulars.

Fig. 11 shows a section of one of the modifications of this range-finder as originally patented in 1904. Two optical squares of the pentagonal type are mounted in a tube T′, at a distance apart which determines the base length of the instrument. At one end of this tube a telescope of the ordinary terrestrial type

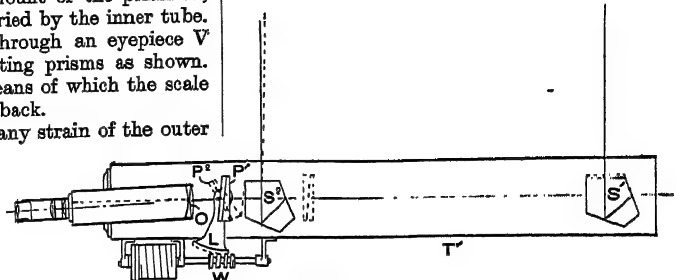


FIG. 11.

is shown. The optical squares are mounted on different levels, so that the light rays from a distant object, passing into the instrument by the reflector S′, enter the upper half of the telescope object-glass O only, whilst light entering by the reflector S² enters the telescope by the lower half of the O.G. In this way, in the example shown, two pictures are projected simultaneously into the focal plane of the eyepiece, which overlap to an extent depending upon the adjustment of the instrument and the distance of the object. The squares being angle-true and adjusted parallel to one another, the images of an object at infinity would overlap and coincide, whereas under the same conditions an object at a short distance would give two images, the separation being dependent upon the range. To bring about coincidence of the two images given in general, two weak achromatic prisms P′ and P² are employed, the first P′ being fixed in position to give minimum deviation to the rays passing through it from the square; whilst the second prism P² is adapted to rotate as shown about an axis parallel to its refracting edge by means of a sector L and worm W. The light reflected by the square S′ passes through the prism P′, and the light reflected by the square S² passes through the prism P². In use, therefore, the worm W, which carries co-axially the range drum D, is rotated to swing the prism P² through the angle necessary to secure coincidence of the two images in the field of view. Thus the range is determined and indicated on the drum D directly.

As an indication of the sensitiveness of this swinging prism micrometer, it may be stated that a prism producing a minimum deviation of 2° has to be swung through an angle of

about 30°, to increase the deviation by 15 minutes of arc. One important advantage claimed for the swinging prism over the sliding prism is that it gives a much more open scale for long ranges. It must, however, be mounted in a parallel beam: in a vergent beam it spoils the definition.

In a later modification of the Cooke 9 ft. naval range-finder, patented in 1920, several features of novelty are introduced. The end optical squares reflect in opposite directions, on to superposed cross mirrors mounted near the middle of the range-finder, from which mirrors the light beams pass, parallel to their original direction, to a single mirror, mounted to direct both beams into a single telescope, which is inclined at an angle of 45° to the axis of the range-finder, but broken by an erecting prism system, which finally directs the light to the eyepiece in such a direction that the observer looks in the direction of the object being ranged upon.

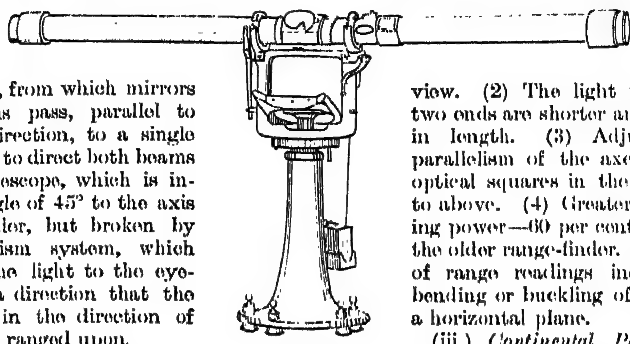


FIG. 12.

Two simultaneously operated swinging coincidence prisms are employed, one behind each of the optical squares. To prevent the intrusion of light from one partial field of view to the other, an ingenious "halving grid" is employed. This consists of a number of stretched threads or cords mounted side by side in a plane containing the axis of the telescope, each thread being at right angles to that axis. This grid practically extends from the object-glass to the eyepiece, thus dividing the telescope into upper and lower halves. To obviate the serious and unsuspected errors which may arise from the fact that the axes (virtual meeting edges) of the end squares are not parallel to one another, and normal to the plane of triangulation, an auto-collimating telescope is mounted within reach of the observer, by means of which a beam of light from a fiducial mark in the eyepiece can be directed to one of the optical squares, by which it is reflected back again to the telescope. Should the axis of the square not be normal, the image of the fiducial mark is seen displaced with respect to the fiducial mark itself. The necessary adjustment having been made, parallelism of the axes is effected simultaneously with any necessary adjustment for halving.

Several advantages are claimed for this modification, as compared with the 1904 range-finder:

(1) Coincidence of the images can be obtained for all ranges across the field of view. (2) The light paths for the two ends are shorter and more equal in length. (3) Adjustments for parallelism of the axes of the two optical squares in the way referred to above. (4) Greater light-gathering power—60 per cent greater than the older range-finder. (5) Accuracy of range readings independent of bending or buckling of the tubes in a horizontal plane.

(iii.) *Continental Patterns.*—The coincidence range-finders as made on the Continent, by Zeiss and others, do not vary substantially in general design from those made in England. A variable-power deflecting prism, mounted in a parallel beam of light, is, however, usually substituted for the travelling prism of the English range-finder, although the latter too is employed.

FIG. 13.

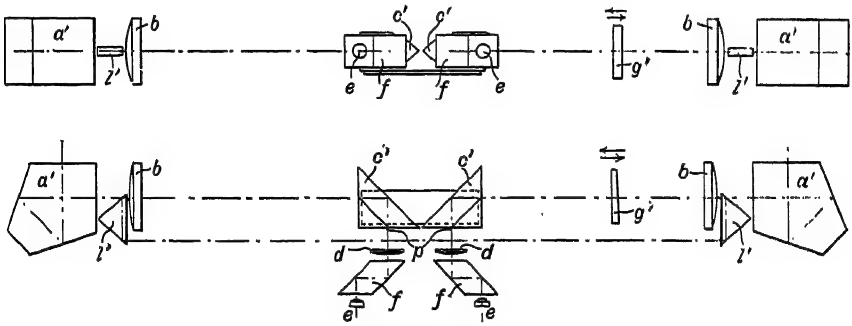
(iv.) *Modern Barr & Stroud Pattern.*—*Fig.* 12 shows the Barr & Stroud 9 ft. base naval range-finder with which the British Fleet was equipped at the battle of Jutland. The range-finder itself, fitted with bearing rings, is adapted to rotate about its own axis in bearings carried by the upper ends of the "Y," which is free to rotate about a vertical axis under the action of a hand-wheel. Adjustment in altitude is effected by means of a radial handle seen on the left of the support.

Finally, *Fig.* 13 represents a monster coincidence range-finder of 100 ft. base which has been constructed by Messrs. Barr & Stroud

for purposes of coast defence. Optically the instrument is practically identical with that of smaller base length. For the purpose of comparison the man in the centre of the figure is shown using the smallest range-finder with a base length of 30 in.

§ (7) Stereoscopic Range-finders, General Construction of.—*Figs.* 14 and 15 show diagrammatically, in plan and elevation respectively, the optical system of a typical stereoscopic range-finder, invented by Carl Zeiss in 1907. The end squares *a′* direct light through telescope object-glasses *b* to right-angled roof prisms *c′*, which deflect the light into the focal plane of the object-glasses, in which two fixed stereoscopic marks *p* occur. These marks are in the focal plane of eyepieces, each of which consists of a pair of lenses *d* and *e*, between which a rhomboidal prism *f* occurs. A ray passing outwards through the eyepiece, left ocular. The image thus produced should superpose on the real mark. It will be noticed that, in this adjusting arrangement, each objective acts as a collimator for its corresponding stereoscopic mark. In operation, therefore, the light from one of the marks is collimated by its corresponding object-glass, from which it passes as a parallel beam to the second object-glass to be brought to a focus in the focal plane of the other mark. These adjusting arrangements will be found dealt with more fully below.

(i.) *Forbes Range-finder.*—One of the pioneers of the stereoscopic range-finder in this country was Professor George Forbes. In 1901 he patented a range-finder constructed of two separate parts, namely, a portable base having a set of reflectors rigidly fixed to it at each end, and a binocular field-glass. In the approved form of the instrument the wander-



Figs. 14 and 15.

therefore, suffers reflections as shown by the surfaces of this rhomboidal prism, with the result that a lateral displacement of the ray occurs. By simultaneous rotation in opposite directions, therefore, of the eyepieces about the axes of the lenses *d*, the separation of the two eye lenses *e* can be adjusted to suit individual observers. It will be noted that this adjustment is secured without any alteration of the distance between the two stereoscopic marks *p*.

In this instrument, which is of the wandering-mark type, the necessary lateral displacement of the picture produced by the right telescope is effected by the axial displacement of a prism *g′*. To detect readily any defect in the infinity adjustment, a subsidiary optical system is employed to project an image of each of the marks *d* into the plane of the other one. Light, say from the mark *p* on the right, is reflected by the prism *c′* to the object-glass *b*, thence by reflection by a prism *l′*, on the right, to a similar prism *l′*, on the left, and finally through the object-glass *b*, on the left, to the prism *c′*, and thence to the mark *d* in the ing-mark was adopted, namely, a fixed line in one eyepiece combining stereoscopically with a similar line in the other eyepiece traversed laterally by means of a micrometer. The marks employed were "pictures of ordinary objects such as a balloon as shown, the ends of the trailing ropes marking the centres of the field." An important advantage was obtained by the use of this balloon mark. The observer naturally brought it over the object being ranged upon, and thus in many cases avoided the psychological difficulty, which so often occurs in stereoscopic range-finding, of being compelled to see the scale, as if behind an object, such as a brick wall, known to be opaque.

In a second specification, taken out in 1903, fixed marks are employed in the focal planes of the eyepieces, and one of the pictures is displaced laterally with respect to the other one, by means of a rotating prism, to bring up the stereoscopic picture into the same plane as that of the image of the two fixed marks. This type, however, as we have seen, was described by Grousilliers.

(ii.) *Zeiss Submarine Range-finder.* — The 3-metre base stereoscopic range-finder made by Carl Zeiss for use in submarines probably represents the highest development of this type of instrument. The range-finder *per se* is contained in a horizontal tube, which forms the horizontal branch of a "T," the vertical branch of which represents a periscope passing into the interior of the boat through the deck. The two pictures of external objects are projected upon stereoscopic graticules, mounted close together and parallel to one another, transverse to the axis of the range-finder, and near its middle. The images in the planes of these graticules, and the range marks, are projected by means of a binocular periscopic system into the focal planes of a pair of oculars at the lower end of the vertical tube. The optical system employed for the range-finder is comparatively simple in character. Light enters by the windows and is reflected along the tube in opposite directions by the usual optical squares, each beam being brought to a focus in the plane of its corresponding graticule by a long focus lens of the telescope object-glass type. These graticules are fixed after being adjusted, and the final stereoscopic picture itself is caused to "wander," *i.e.* move backwards and forwards in the line of sight, for ranging. In one half of the tube a weak refracting prism is mounted to slide backwards and forwards, along the axis of the tube, between the object-glass and the graticule, for the purpose of displacing laterally the image seen in the corresponding ocular, for the purpose of bringing about that adjustment, or wandering of the stereoscopic image in the line of sight, necessary to bring it up to the fiducial mark—the crossing point of the two linear stereoscopic scales. In the other half of the range-finder, a plane-parallel plate is mounted between the object-glass and the graticule, as so to tilt about a horizontal and transverse axis, for the purpose of raising and lowering the picture in the corresponding ocular, to bring it to the same vertical height as the picture with which it has to be stereoscopically fused.

Direct observation of the graticules being impossible, the range-finder is combined with a binocular periscopic system, with the aid of which the graticules can be observed from a position inside the submarine. For this purpose a pair of optical squares are mounted between the two graticules to face in opposite directions and deflect the beams of light passing through the graticules at right angles to the range-finder, and in a horizontal plane, each beam being again reflected vertically downwards, through a collimating lens, and into a telescope by which, finally, an image of the corresponding graticule is presented to the observer. This periscopic system is duplicated for the second half of the range-finder.

§ (8) Height-finders.—In the early days of the war the need for a height-finder as an auxiliary to anti-aircraft guns was recognised. Many tentative solutions of the problem were proposed from time to time, but the final solution, resulting in the provision of an efficient short-base service instrument, was due to Messrs. Barr & Stroud in 1917.

In ranging on an aeroplane, a horizontal range-finder is rotated about its axis to elevate the line of sight and bring the aeroplane into the field of view. Coincidence is then effected and the range determined. It will be observed that in this operation we have got the two necessary elements for solving the right-angle triangle defined by the observer, the target (aeroplane), and a point on the ground immediately below the target. The range gives the hypothenuse of this triangle, and the angle of elevation, or sight, one of the acute angles. With these two elements, it is a simple slide-rule problem to determine either the height of the aeroplane above the ground, or its horizontal range, that is, the distance from the observer to the bottom of the perpendicular dropped from the aeroplane. This was the principle upon which some of the first height-finders were constructed. One of the scales of the range-finder was operated more or less automatically by the motion in altitude of the range-finder, to set off the logarithm of the sine of the angle of sight, and then, by a hand-setting of the range, the equation, log H = log R + log sine angle, was solved and the height H determined. All these arrangements, however, gave the range primarily, so that so long as the range of the target was altering constant adjustment of the working head was necessary. What was wanted was an instrument the constant-coincidence surface of which was not a hemisphere, as in the range-finder, but a horizontal plane—a height-finder.

(i.) *Barr & Stroud.*—In Barr & Stroud's specifications the invention described gives both range and height automatically. It consists of a known range-finder of the coincidence type so modified as to function also as a height-finder. This in effect is carried out by fitting the range-finder with an automatically operated slide rule. The two sliding parts are constrained to move, one proportionately to the logarithm of the range, and the other proportionately to the logarithm of the sine of the angle of sight. The height can thus be read off directly. The motion of the adjusting head of the range-finder and the elevation of the range-finder are transmitted to the sliding scales through the medium of logarithmic toothed gearing.

The working head is transferred from the range-finder to the height-indicating attach-

ment, so that the instrument acts primarily as a height-finder. No manipulation of the working head, therefore, is necessary, so long as the target under observation retains its height.

Fig. 16 shows the range-finder, with a base length of two metres, mounted on a tripod

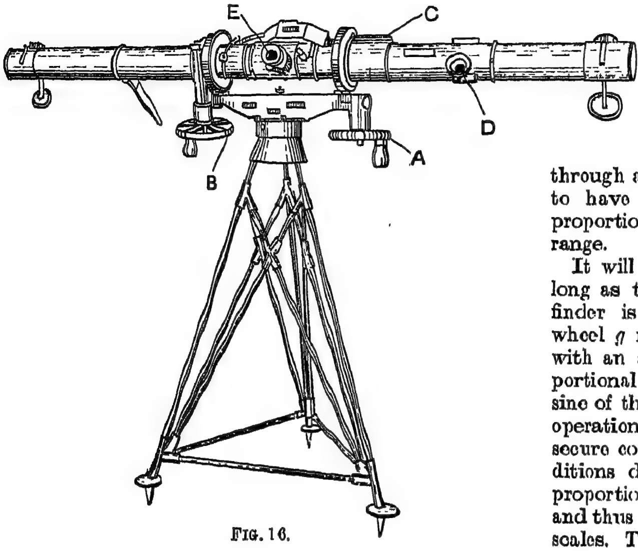


FIG. 16.

stand, upon which it is rotated about a vertical axis by the operation of a two-speed hand-wheel A. The operation of a second hand-wheel B rotates the range-finder about its own axis to elevate the line of sight. Coincidence is maintained by the working head C, whilst D is a finder.

Three observers are required: the first looks into the eyepiece E and operates, as required for the maintenance of coincidence, the working head C, and the hand-wheel B to direct the line of sight in altitude. The second observer uses the finder D and is responsible for the training of the range-finder by means of the hand-wheel A. The third observer reads off the height- and range-scales.

The instrument is essentially a range-finder with a gearing attachment, as shown by *Fig.* 17, which automatically, as the target is kept in the field of view and coincidence maintained, indicates the height and range. The end of the shaft of the working head *a* carries a pinion *e*, which gears with a bevel-wheel *f*, carrying the height-scale, and also—on a radial shaft—the jockey wheel *d*, which gears with the upper and lower members of a differential pair of spur wheels *g* and *h*. The upper member *g* of this pair is driven directly from the elevating head *b*, through logarithmic spiral gear which gives to *g* an angular displacement proportional to the logarithm of the sine of the angle of sight. The lower member *h* of the differential pair carries the range-scale, and is geared up to the deflecting range-prism through a clutch, not shown, so as to have an angular displacement proportional to the logarithm of the range.

It will be seen, then, that so long as the elevation of the range-finder is not changed the upper wheel *g* remains fixed in a position with an angular displacement proportional to the logarithm of the sine of the angle of elevation. The operation of the working head, to secure coincidence, under these conditions drives simultaneously and proportionately the wheels *f* and *h*, and thus also the height- and range-scales. These conditions are realised by ranging on a target flying away from the observer along a straight line. With the working head fixed, the elevation of the instrument operates the upper wheel *g*, and also, through the jockey wheel *d*, the lower wheel *h*, and thus the deflecting prism and

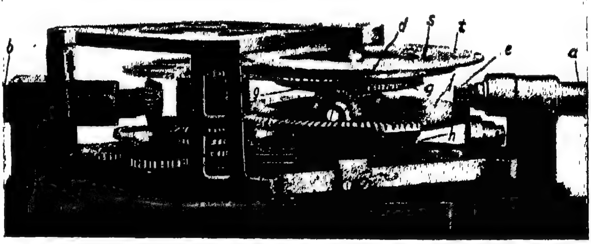


FIG. 17.

the range-scale. Since, however, the wheel *f* remains fixed, the height-scale is unaffected. This represents the conditions of use when the target ranged upon is moving about in a horizontal plane.

(ii.) *Rottenburg and Willan.*—An ingenious optical solution of the height-finder problem was patented by Rottenburg and Willan in 1916. By this invention a second prism is combined with the deflecting prism, and the

two prisms rotated equally and oppositely, about the axis of the instrument, by the adjustment of the instrument in altitude necessary to keep a target in the field of view. The result of this arrangement is that whereas, in the ordinary Barr & Stroud range-finder, a simple deflecting prism of constant power is employed, in Rottenburg and Willan's invention a variable-power coincidence prism is employed, the power of which is such a function of the elevation, or angle of sight, that the instrument is adapted to operate as a height-finder; no operation of the working head being necessary so long as the target remains at the same height.

II. Optical Elements of the Range-finder

§ (9) End Reflectors: Optical Squares.—In the early days of the range-finder, plane-reflectors, or right-angle total-reflecting prisms, were employed as end reflectors, but these forms were open to the serious objection that any bending of the frame of the range-finder in a horizontal plane, due to unequal heating, for instance, caused these end reflectors to turn in effect about a vertical axis and thus alter the angle through which the rays were deflected; and it was not until these reflectors were replaced by optical squares that the difficulty was overcome.

When a ray of light is reflected in succession by a pair of plane reflecting surfaces, as in the case of the sextant, the angle through which the ray is deviated is equal to twice the angle included between the reflectors, so long as the reflections take place in a plane at right angles to the meeting edge of the reflectors. If then these reflectors are inclined at an angle of 45° to one another, the deviation produced by successive reflections is equal to a right angle, and we have an optical square; that is, an optical device which gives us an angle of deviation which is independent of the angle at which the entering ray strikes the first reflecting surface. The end reflectors of small range-finders are now, almost invariably, optical squares made from a single pentagonal-shaped block of glass. In use, any given ray passes through the entrance face practically normally, and after internal reflections at the two bounding surfaces meeting at an angle of 45°, passes through the exit face, again practically normally, with a deviation of a right angle which is independent of any small rotational displacement of the prism in its own plane. The conditions of accuracy for such a square are (1) homogeneous glass; (2) reflecting surfaces inclined at an angle of 45°; and (3) plane entrance and exit faces, meeting under an angle of 90° in an edge parallel to the meeting edge of the two reflecting surfaces.

Such squares in large sizes become impracticable, mainly on account of the loss of light incurred by absorption, so that considerable attention has been devoted of late years to the designing of optical squares comprising two plane reflectors, the included angle of which shall be maintained with great accuracy, in spite of bending stresses and temperature changes. Mr. Dennis Taylor has paid particular attention to this subject, as will be seen by reference to his various patents.

§ (10) Deflecting Systems for Coincidence Range-finders.—The following optical micrometrical devices for effecting coincidence have been used, but the first three are now those commonly employed:

(*a*) A longitudinally travelling prism in a convergent beam, *i.e.* between the telescope O.G. and the separating prism (Barr & Stroud and others).

(*b*) A tilting prism in a parallel beam (Taylor).

(*c*) Oppositely rotating prisms in a parallel beam, acting as a variable-power prism (Zeiss).

(*d*) Rotary end reflectors.

(*e*) Laterally displaced object-glasses.

(*f*) A tilting plane-parallel plate of glass in a convergent beam.

(*g*) Differently magnified images brought into coincidence by the rotation of the range-finder about a vertical axis (Eppenstein).

The first of these systems has the great merit of simplicity, and less liability, therefore, to derangement. The prism itself, as has already been pointed out, carries an ivory scale some seven or eight inches in length, graduated in ranges which are read off opposite a special eyepiece provided for the purpose. In the second system the advantage of a more open scale at long ranges is secured. In the third system two weak prisms are mounted, face to face, between the end squares and the O.G.'s, so as to rotate equally and oppositely. In this way a variable deflecting prism is obtained which produces deflections in a horizontal plane only. This system is open to the defect that the maximum deflecting power of the combination is obtained by a rotation through 180 degrees only, of the prisms. This difficulty has been, to some extent, overcome in the Zeiss range-finder by the introduction of a multiplying system of gear wheels, which gives to a circular range-scale an angular displacement practically twice that of the deflecting prisms.

In the ordinary coincidence range-finder it is important that the two images brought into coincidence should be magnified equally, otherwise different range-readings will be obtained according as the coincidence happens to be made near the middle or towards the ends of the separating line in the field of view.

In a new type of range-finder, introduced

within the last few years by the firm of Zeiss, and invented by Eppenstein, the two images are differently magnified; so that as the range-finder is rotated about a vertical axis, the more magnified image moves more rapidly across the field of view than the less magnified image does—starting behind it, it may be, catching it up, and then passing it. From this it follows that for every range between infinity and the minimum range required, there is one place, and one place only, along the separating line in the field of view, at which coincidence of the two images can occur. A range-scale is therefore drawn along the separating line, extending from infinity on the left to the minimum range required on the right. In ranging on an object, therefore, no adjustment is required beyond that of rotating the range-finder about the vertical axis until the two images of the object being ranged upon are seen to be in coincidence.[1]

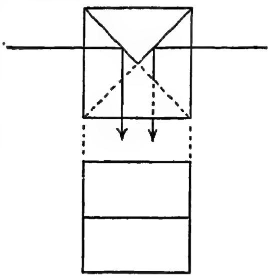


FIG. 18.

§ (11) ASTIGMATISERS.—When the object being ranged upon is ill-defined and illuminated in patches, or a point such as a searchlight,

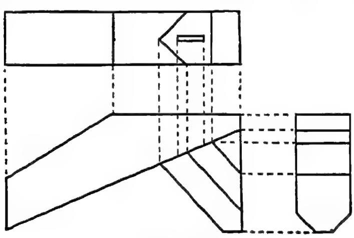


FIG. 20.

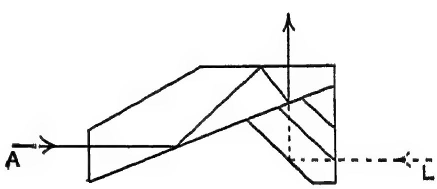


FIG. 21.

astigmatising systems are interposed in the ray paths on both sides of the instrument. This system may consist of a pair of negative cylindrical lenses, with axes horizontal, which can be thrown at will into the ray paths. The action of these lenses is to draw out the image in the eyepiece into vertical streaks across the separating line, when coincidence can be effected in the usual way.

§ (12) SEPARATING PRISMS.—The separating prism combination is perhaps the most important optical element in the range-finder It is situated between the object-glass of each of the telescopes and its focal plane; its function being to deflect the imaging cones of rays, which it receives from the ends of the range-finder, at right angles into the common focal plane of the eyepiece and the two objectives, with fine and sharp boundary lines between the partial fields, and at the same time giving to the images in these fields the desired orientation. An ideal separating prism would give a field of view with an invisible dividing line when directed to a clear sky. In early forms of range-finders the separating prism consisted simply of two rectangular prisms superposed as shown by *Fig.* 18, the light from the left window producing an image in the upper part of the field, and the light from the right window producing similarly the image in the lower part of the field. Another simple arrangement is shown by *Fig.* 19. The two prisms are placed edge to edge as shown, with the result that the field is divided by a vertical line instead of a horizontal one. In both these cases, however carefully the prisms may be made, the dividing line is shown to be very rough and badly illuminated when viewed, as it is, in the range-finder, with an eyepiece magnifying from 10 to 12 times.

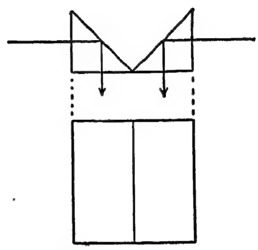


FIG. 19.

In the separating prism system employed in some forms of the Zeiss range-finder (see *Figs.* 20, 21, and 22), two prisms are cemented together with the surface of contact silvered, except for a small rectangular opening shown in plan view (*Fig.* 20). This rectangular opening in the silvering occurs in the common focal plane of the eyepiece and the telescope object-glasses. Light from the right eye of the range-finder is totally reflected from two faces of one of the component prisms, as shown by *Fig.* 21, and then from the silvered inter-face into the eye of the observer. Any light, how-

[1] For information on the calibration of the range-scales for instruments of different types see below, § (16), etc.

ever, which falls upon the unsilvered rectangular opening passes through it and away. Light from the left eye of the range-finder, after being reflected from the roof faces of the second component prism (see *Fig.* 21), passes vertically upward until it reaches the inner face, at which part is stopped by the silvering and part passes through the rectangular opening to the eye of the observer. The picture in the field of view, therefore, is for the most part produced by light from the right eye of the range-finder, but the narrow rectangular opening seen in the middle of this field of view frames a picture due to light from the left eye only of the range-finder. A small subsidiary prism cemented into the angle on the right, between the two main component prisms already referred to, serves to reflect the range-scale reading into the field of view. It will be seen that since the small rectangular opening in which one of the partial pictures appears is formed on a prism face, which is not at right angles to the optical axis of the eyepiece, it cannot be

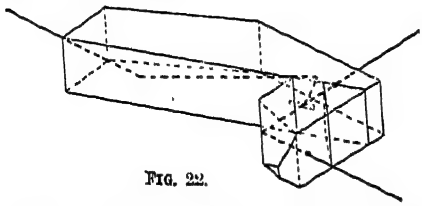


Fig. 22.

focussed up simultaneously throughout its entire length. This is a defect.

Fig. 23 shows, in side elevation and plan respectively, one of the various forms of eyepiece prism systems used by Messrs. Barr & Stroud in their range-finders. The top prism receives the light from the left eye of the range-finder, and the bottom prism acts similarly for the right eye of the range-finder. The ray L, shown as a full line, enters the top prism and is reflected at 1, at right angles in a forward direction—that is, away from the observer—and in the horizontal plane. It is again reflected at right angles into a downward direction at 2, ultimately being reflected at 3 to the separating prism A; the horizontal dihedral edge of which, formed by the meeting of the two prism faces at an obtuse angle, is in the common focal plane of the eyepiece and of the telescope objective, and forms the fine separating line in the field of view. All rays from the left eye of the range-finder fall upon the two faces of the separating prism, but it is only those falling upon the lower face which are refracted through the opening in the diaphragm to the eye of the observer. Those rays which fall upon the upper face are refracted as shown by the long-and-short dotted line so as to be stopped by the diaphragm. Rays reaching the separating prisms from the right eye of the range-finder are dealt with in a similar way, but, in this case, the rays falling upon the upper face of the prism are the only ones which ultimately reach the eye of the observer—those falling upon the lower face are refracted so as to be stopped by the diaphragm. In this way the separating prism acts to divide the field of view into two parts, the upper one of which receives light from the right eye only, whilst the lower part of the field receives light from the left eye only, the two partial fields of view being separated by a fine sharp line. A compound cylindrical lens B can be thrown

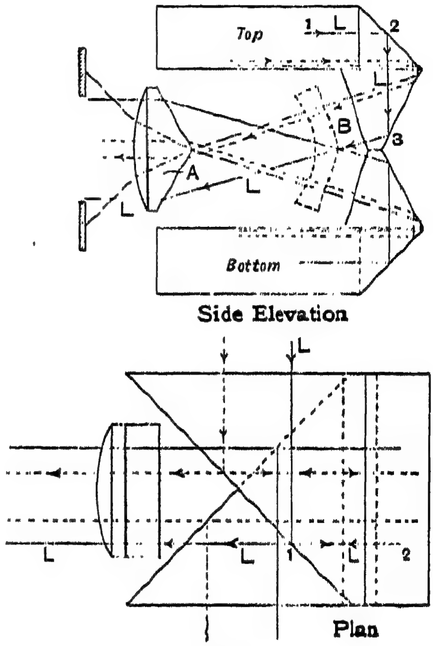


Fig. 23.

into the ray paths at will to astigmatise the final observed image.

§ (13) Stereoscopic Graticules.—In the Zeiss instruments, each graticule of the pair used in instruments of the fixed-scale type has a number of range-marks, which are numbered at intervals and arranged along a zigzag line. When these scales are combined visually, a single stereoscopic range-scale is seen, which, starting in the foreground of the picture, extends away from the observer indefinitely.

For the wandering-mark type a pair of similar marks, one in each graticule, is employed. In later instruments, however, especially those of the wandering-picture type, each of the graticules bears a number of range-marks (arrow-heads pointing downwards) arranged in two straight lines, approximately horizontal, but crossing one another at the

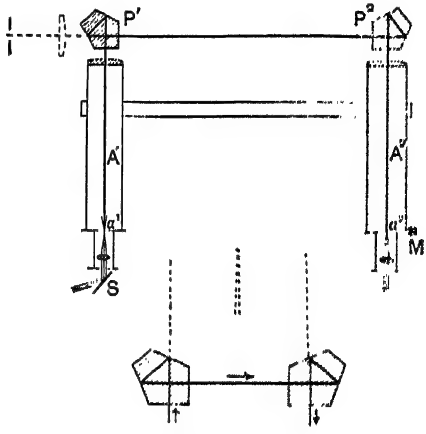
middle, at an angle of about 10°. These scales, when visually combined, present to the eye two overhead linear range-scales, each starting near at hand and receding into the distance; but whilst one is seen bearing away to the right, the other appears to bear away to the left, so that at a mean distance they appear to cross, the crossing point having a common range-mark. In observing with this type of range-finder, the deflecting prism is adjusted until the object being observed is seen to be at the same distance away as the range-mark common to the two scales at their crossing point. This design of graticule was invented by Zeiss about 1908. Greater accuracy when ranging on isolated objects is claimed for it over that obtained by the use of two single marks only.

§ (14) Range-scale Adjusting Apparatus for Range-finders.—In an instrument like the range-finder, which, under the severe conditions of use, is called upon to work at such a high order of accuracy, it will be readily understood that the maintenance of the various parts of the instrument in correct adjustment is a very severe problem for the designer and manufacturer. So far as the range-scale is concerned, the instrument can always be tested and adjusted when very distant or celestial objects are available. It is then only necessary to direct the instrument to a distant object—the moon, say—and operate the working head to effect coincidence of the two images. The range-scale should then read "infinity." If it does not do so, then the range-scale is set to indicate "infinity" and the adjusting prism forming one of the windows, above referred to, is rotated until coincidence is effected.

(i.) *Lath Adjuster.*—In the simplest form of apparatus for effecting this infinity adjustment, apart from ranging on real and distant objects, two vertical lines are drawn upon a lath or board at a distance apart equal to the base length of the range-finder with which it is to be used. When this board is fixed up at a distance of some hundreds of yards from the range-finder, and parallel to it, then one line is projected into one partial field of the eyepiece and the other line is similarly projected into the other. If the range-finder be in correct adjustment the range indicated will be "infinity" when coincidence has been effected. Quite early, however, in the history of the range-finder, it was felt that this method was too crude and cumbersome.

(ii.) *Abbe's Method.*—In the year 1893 Professor Abbe devised apparatus for correcting and adjusting telemeters or range-finders, which was based upon optical means which have practically been common to all the later and more developed apparatus for securing the same end.

Abbe's apparatus was particularly applicable to range-finders of the type (stereoscopic) in which two independent telescopes are employed. In *Figs.* 24 and 24A, for example, the invention is shown diagrammatically as applied to a stereoscopic range-finder of the wandering-mark type, that is, the range is determined by the adjustment of the micrometer M, which is necessary to give to the observer the impression that the wandering stereoscopic mark seen in the field of view is at the same distance as the object being ranged upon. The adjusting apparatus consists of a pair of pentagonal prisms P′ and P², mounted in front of the two telescope object-glasses as shown. The left eye of the observer



Figs. 24 and 24A

is then taken away from the telescope A′, and a mirror S is positioned to illuminate the fixed stereoscopic scale mark a'. Light from this mark, therefore, issues from the object glass as a collimated beam and enters the prism P′, by which it is reflected at right angles to the prism P², and then similarly into the second telescope A², by which it is brought to a focus in the same plane as the wandering-mark a^2. If the range-finder is in correct adjustment, and the prisms P′ and P² reflect the beam of collimated rays accurately through 180°, then it follows that the image of a' will be projected on to the wandering-mark a^2, when the micrometer M is set to infinity. Should this not be the case, the instrument is adjusted until it is so. The great advantage of this arrangement is that each of the pentagonal prisms acts as an optical square, that is, the deviation produced by it is practically constant and, further, independent of any small rotation of the prism about an axis normal to the plane of the diagram. When the base length of the range-finder is small, the pair of optical squares may be made from a single piece of glass.

The prisms P′ and P^2 may together produce a deviation of the collimated rays something less than 180°. Let this angular defect from $180° = \theta$, then the rays will enter the telescope A^2, as they would from a real object on the axis of the telescope A′, and at a distance equal to that at which the base of the instrument subtends the angle θ. If this angle θ and its corresponding range are known, then the testing operation consists as before in projecting the mark a' on to the mark a^2, but the micrometer M should, when this adjustment has been made, indicate the corresponding range.

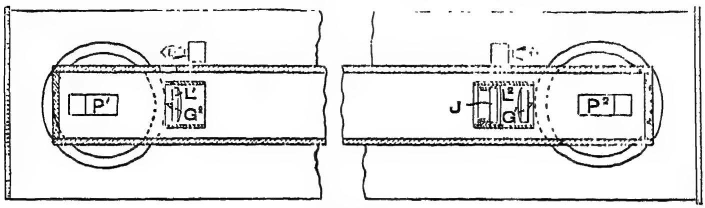


FIG. 25.

In a later modification of Abbe's apparatus, invented by König in 1904–1908, the necessity for using one of the telescopes of the instrument as a collimator is avoided by fitting a collimator independently.

(iii.) *Barr & Stroud's Method.*—An important instrument for adjusting range-finders was brought out by Messrs. Barr & Stroud in 1906. The optical principle upon which this instrument was constructed is shown by *Fig.* 25. Fiducial marks G′ and G^2 are borne by lenses L^2 and L′ respectively, of equal focal length and separated by a distance equal to that focal length. Each lens as shown, therefore, acts as a collimator to the fiducial mark on the other lens, with the result that the emergent beams are collinear, a property which is independent of any small lateral displacement of either of the lenses L′ and L^2, provided that the marks G′ and G^2 are mounted at the principal points of the two lenses. These collinear beams enter prisms P^2 and P′ respectively, from which they are reflected parallel to one another, and are therefore in a condition to enter the range-finder and give images of the lines G′ and G^2, equivalent to the images of a single object at infinity.

Fig. 26 shows in plan the application of this instrument to a range-finder, and in addition an adjusting prism J, which is adapted to rotate about an axis coincident with that of the two lenses G′ and G^2. By the rotation of this prism, therefore, the two issuing beams of light can be made either parallel to one another, or to diverge at the angle corresponding to that between beams of light entering the range-finder from an object at a known distance. The adjusting apparatus may, as shown, be made separate from the range-finder, or it may be attached to the framework of the range-finder itself.

§ (15) HALVING ADJUSTMENT.—In general the picture seen in one of the fields-of-view of a range-finder is duplicated and inverted in the second field, but it may happen that corresponding image points do not occur at equal distances from the separating line, so that when the infinity adjustment has been made, the image in the upper field is not a simple reflection of that in the lower field, with the line of separation as a line of symmetry.

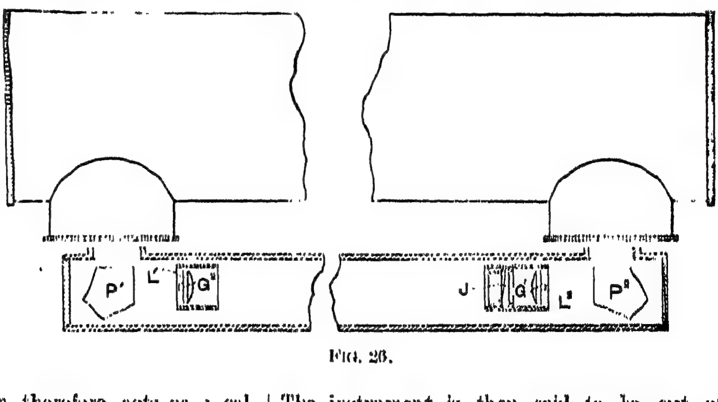


FIG. 26.

The instrument is then said to be out of halving adjustment. This want of adjustment, so long as straight edges at right angles to the separating line are being ranged upon, is not of great importance, but when a line inclined to the separating line is used for ranging purposes a very serious error is introduced by this want of adjustment. To effect the necessary adjustment it is clear that one picture must be displaced in the field-of-view vertically with respect to the other one. This may be carried out in the range-finder

by (1) tilting one of the pentagonal reflectors about a horizontal axis inclined at an angle of 45 degrees to the axis of the tube and parallel to the path of the normally incident ray as it passes from one reflecting surface to the other; (2) by displacing one of the telescope object-glasses in a vertical direction; (3) by the rotation in its own plane of a weak prism originally mounted with its edge vertical; or (4) by tilting a thick plane-parallel plate in a convergent beam of light.

The halving adjustment having been made, the correction of the infinity adjustment is made by shifting the upper field in a direction parallel to the dividing line as already described.

III. Calibration of Range-scales

§ (16) Coincidence Range-finders. (i.) *Sliding-prism Type.*—The relations which must exist in a coincidence range-finder between the length of its base b, the focal length f of its objectives, the deviating power δ in circular measure of its sliding deflecting prism, and the length L of its range-scale between the infinity mark and the mark for the minimum range R_m required, can be obtained as follows:

Imagine a range-finder adjusted on an object at infinity, and that the object then moves up along the straight line joining it with the left eye of the range-finder till its distance from the range-finder is equal to the minimum range R_m which it is desired to indicate on the range-scale. As the object thus moves, the parallax angle changes from 0 at infinity to b/R_m at the minimum range, and the image projected by the right half of the range-finder into the focal plane of the eyepiece moves out of coincidence through a distance d such that

$$\frac{d}{f}=\frac{b}{R_m}. \quad . \quad . \quad . \quad . \quad \text{(A)}$$

The image projected by the left half of the range-finder under the conditions named would remain stationary. To re-establish coincidence of the images the sliding prism with angular deviating power δ must be moved through a distance L, along the axis of the beam of light, such that $d=L\delta$; so that substituting this value of d, in equation (A) above, we obtain

$$L\delta=\frac{f.b}{R_m}. \quad . \quad . \quad . \quad \text{(B)}$$

To apply this formula, suppose it is required to find the length L, when the base length of the range-finder is 1 metre (39·37 inches), the focal length f is 11 inches, the minimum range required is 250 yards, and the deviating power of the prism is one unit in a distance of 170 units (equivalent to 20 minutes of arc). Then reducing the linear dimensions to inches in equation (B) we find that

$$L=\frac{11\times1\times39{\cdot}37\times170}{250\times36}=8{\cdot}18 \text{ in.}$$

(ii.) *Rotating-prism Type.*—In range-finders, such as the Zeiss, in which coincidence is effected by imparting equal rotations in opposite directions to a pair of similar prisms mounted in the parallel beam between the end reflector and the telescope O.G., it is desirable to secure the maximum possible openness of the range-scale by giving to each of the prisms an angular deviating power δ, such that a rotation of these prisms through 180°, *i.e.* from the position shown in plan by *Fig.* 27 to the position shown by *Fig.* 28, should allow of coincidence for all ranges between infinity and the desired minimum range R_m.

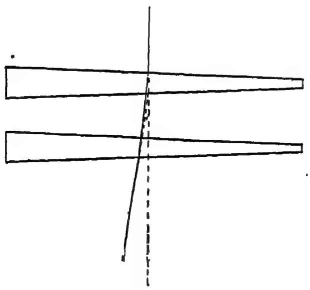


Fig. 27.

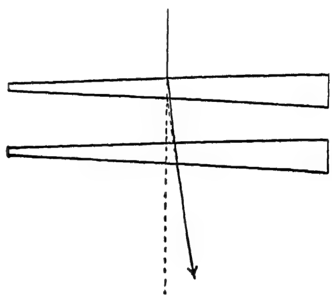


Fig. 28.

The angular deviating power of each prism being δ, a rotation through 180° will effect a deviation of any ray in the horizontal plane through an angle equal to 4δ, and the deviation δ_1, produced by a rotation through any angle α less than 180°, as shown by *Fig.* 29, is obtained from the equation

$$\delta_1=2\delta(1-\cos\alpha), \quad . \quad . \quad . \quad \text{(C)}$$

it being remembered that when α is between 90° and 180° the cosine is negative in value.

Thus for rotations of 0°, 90°, and 180° the deviations are 0, 2δ, and 4δ respectively. The angle 4δ must therefore be made equal to the

maximum angle of parallax b/R_m to be measured, or

$$\delta = \frac{b}{4R_m}. \quad . \quad . \quad . \quad . \quad (D)$$

Thus when $b=1$ metre and $R_m = 250$ m.,

$$\delta = \frac{1}{1000},$$

approximately $3\frac{1}{2}$ minutes of arc. It will be noticed that the focal length of the O.G. does

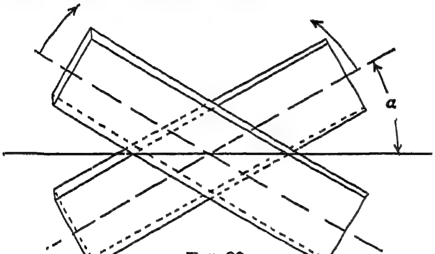


FIG. 29.

not enter into equation (D), for the reason that the deviation is effected in the parallel beam and not in the convergent beam, as in the case of the sliding-prism type of range-finder.

For any given range R the parallax angle $\theta = b/R$, and since δ must be equal to θ we have, by substituting in equation (C) above,

$$\frac{b}{R} = 2\delta(1 - \cos\alpha),$$

or

$$\cos\alpha = 1 - \frac{b}{2\delta R}, \quad . \quad . \quad . \quad (E)$$

where α is the angle of rotation of one of the deviating prisms from its infinity position necessary for bringing about coincidence. By giving to R in the above equation the successive numerical values it is desired to indicate on the range-scale, the corresponding angle α can be found.

Thus in the case of the range-finder considered above, with a base length of 1 metre and deflecting prisms each with an angular deflecting power of one in a thousand, equation (E) becomes

$$\cos\alpha = 1 - \frac{500}{R},$$

from which the following table has been calculated as an example:

Range (metres).	Angle.	Range (metres).	Angle.
20,000	12° 50′	800	67° 59′
10,000	18° 11′	700	73° 23′
5,000	25° 50′	600	80° 22′
4,000	28° 58′	500	90°
3,000	33° 35′	400	104° 29′
2,000	41° 25′	300	131° 49′
1,000	60°	250	180°
900	63° 38′	..	..

These angles must, of course, be multiplied by the gearing ratio adopted—usually two.

(iii.) *Two-magnification Type.*—In this type of range-finder an eyepiece range-scale, extending along the lower side of the coincidence line, is graduated so that when the instrument is in adjustment the coincidence of the two differently magnified images of an infinitely distant object can only be obtained at the place along the coincidence line at which the infinity mark of the scale occurs. Let *Fig.* 30 represent the field-of-view of such a range-finder, and suppose that the images of an infinitely distant object have been brought into coincidence on the line marked "∞." Now suppose the object to move up along the straight line, joining it with the left eye

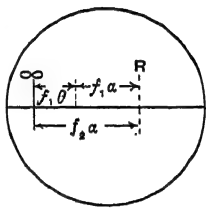


FIG. 30.

of the range-finder, to a distance at which the base length b of the range-finder subtends an angle θ. Then the partial image, in the upper part of the field-of-view, produced by light passing through the right eye of the range-finder, will, during the supposed movement of the object, move from left to right through a distance $f_1\theta$, where f_1 equals the focal length of the corresponding telescope objective, whilst the partial image due to the left eye, in the lower part of the field, remains on the infinity mark. Now suppose the range-finder to be rotated through a small angle α about a vertical axis, in the counter-clockwise direction as seen from above, both the images will then move to the right, but the more magnified image, in the lower part of the field, will move more quickly than the one in the upper part of the field, and eventually catch it up at a distance d from the "∞" mark, such that $d = f_2\alpha$, where f_2 is the focal length of the telescope objective on the left. Since the image in the lower field is more magnified than the image in the upper field, it follows that $f_2 > f_1$, and that in consequence, whilst the lower image moves through the distance $f_2\alpha$, the upper image will move through a shorter distance $f_1\alpha$. Thus

$$d = f_1\theta + f_1\alpha, \quad . \quad . \quad . \quad . \quad (1)$$

but

$$d = f_2\alpha, \text{ i.e. } \alpha = \frac{d}{f_2}, \quad . \quad . \quad . \quad (2)$$

and

$$\theta = \frac{b}{R}. \quad . \quad . \quad . \quad . \quad (3)$$

Substituting these values of α and θ in equation (1) we get

$$d = \frac{f_1 b}{R} + \frac{f_1 d}{f_2},$$

and solving for d we obtain

$$d = \frac{b f_2 f_1}{R(f_2 - f_1)} \quad . \quad . \quad . \quad (4)$$

which shows that the eyepiece range-scale is a reciprocal one, since b, f_1, and f_2 are constants for the range-finder. Equation (4) may be written

$$\frac{1}{f_1}-\frac{1}{f_2}=\frac{b}{d\,.\,R}\text{ (a constant)}, \qquad (5)$$

i.e. the difference in the focal powers of the two objective systems must be made equal to the base length divided by the product of the length of the range-scale, and the range corresponding to that length.

In practice the total length of the range-scale is limited to about 1 cm., and the minimum range required is, say, 250 metres, so that if the base length equals 1 metre, then

$$\frac{1}{f_1}-\frac{1}{f_2}=0{\cdot}4,$$

the focal lengths being expressed in metres. From this equation, one focal length being assumed, the other can be found. Thus when $f_1=25$ cm., f_2 must be made equal to 27·7 cm. In a range-finder of this type, made by Bausch and Lomb, $f_1=16{\cdot}26$ cm., $f_2=16{\cdot}78$ cm., $b=0{\cdot}8$ m., and $d=0{\cdot}95$ cm. for the minimum range of 450 m.

§ (17) Stereoscopic Range-finders: Fixed Scale and Wandering-mark.—In these types of range-finders (see *Fig.* 8) the parallactic displacement D, in the common focal plane of the two objectives, is obtained from $d=bf/R$, where, as before, B is the base length of the range-finder, f the focal length of the telescope objectives, and R the range. The distance between the two infinity marks —marks upon which the two images of a star would be simultaneously projected—in the common focal plane of the objectives may, by suitable disposition of the reflectors, be given any value, usually that of the average interpupillary distance of the men likely to use the range-finder. This distance, once determined and adjusted, remains fixed, that is, it is not affected by any adjustment of the interocular distance necessary for individual observers. A fixed scale might consist simply of a single mark in one ocular field and a number of marks in the other field, corresponding to the series of ranges for which the instrument is to be graduated. With such an arrangement as this the image of the single mark, in the use of the instrument, would be fused stereoscopically in succession with each of the marks in the other ocular field. In practice, however, it has been found better to provide for each of the ranges to be indicated a separate pair of marks, one in each ocular field, but the principle of the calibration is the same. Thus for a range-finder in which the product bf is equal to k, say, and in which the distance between the infinity marks is equal to A, say, the distance apart of the pair of marks for any finite range R_1 must be made equal to $A-k/R_1$.

The same principles apply in the case of the wandering-mark. Here one ocular field contains a fixed mark, whilst in the second field a second mark is adjusted transversely to the line of sight by means of a micrometer screw, the drum of which is calibrated for the ranges corresponding to the parallactic displacements required for the given ranges according to the equation above.

§ (18) Conclusion.—From the story now told, it will be seen that whilst the English naval and military authorities accept the coincidence type of range-finder as the most efficient service instrument, the German authorities on the other hand, under the guiding and compelling genius of Carl Zeiss, have favoured the stereoscopic type. Now in the case of a coincidence instrument there is no question but that a comparatively large percentage of service men can be trained to work it efficiently, whilst in the case of the stereoscopic instrument the conditions are very different. English military opinion tends to the conclusion that only about 5 per cent of men tried ever become efficient range-takers with the stereoscopic instrument, and this conclusion is supported by German experience. In the German training-schools we believe that, of the total number of the men tried, only about 15 per cent are finally selected for training. It is claimed that of these men a certain very small number, after an extended period of trial, attain a stereoscopic differentiating power of 4 seconds of arc. This compares favourably with the standard reached by the best range-takers with the coincidence type. The average, however, claimed by the Germans is about 10 seconds—should a man not be able to attain an accuracy of 20 seconds or less, he is rejected.

It must be remembered that the remarkable accuracy which has been claimed for both types of range-finder has only been secured by tests made under the most favourable conditions—laboratory conditions, in fact—and it may be that, as regards their best performances under those conditions, the one type can claim no decided advantage over the other. Tests under service conditions, however, are the only ones which have any useful discriminating value, and these are not so easily carried out. A range-taker, under the intense nerve strain brought about by a modern battle, may find that he can obtain very much better results with one type of range-finder than the other, and, so far as evidence is available, there can be little doubt that under the severest service conditions the coincidence has a very decided advantage over the stereoscopic type.

Now the battle of Jutland was fought with 9-feet Barr & Stroud coincidence range-finders on the one side, and 3-metre Zeiss stereoscopic range-finders on the other. Very few instruments of greater dimensions than these were used by either side. It is, however, strange to say, very difficult to arrive at a just conclusion as to the relative merits of the work done with these instruments. Each side has shown a desire to give full credit to the work of the other. When, however, everything is taken into consideration, there appears to be very little doubt that the days of the stereoscopic range-finder, as a chief service range-finder, are numbered. This is not to be wondered at when it is remembered that, before the war, Germany had held out against the world in adopting the type. No doubt it will still be made and used for special work, possibly against air-craft. Any one who has experienced the ease and precision with which a stereoscopic range-finder can be directed to and ranged on a mere wisp of floating cloud will readily realise that for work of this kind it stands out unrivalled.

LIST OF THE MORE IMPORTANT BRITISH PATENTS FOR RANGE-FINDERS[1]

Specification No.	Year of Application.	Name.	Subject-matter.
357	1860	Adie	Coincidence: tilting mirror type
608	1863	"	Coincidence: tilting mirror type
8043	1885	Mallock	Coincidence: tilting mirror type
12404	1886	Christie	Coincidence: lateral displacement of O.G.
9520	1888	Barr & Stroud	Coincidence: travelling deflecting prism, etc., etc.
13507	1893	" "	Coincidence: astigmatisers, separating prisms
17048	1893	Grousilliers	Stereoscopic
17868	1893	Abbe	Infinity adjusters
3172	1901	Barr & Stroud	Two objectives side by side
5267	1901	Forbes	Stereoscopic
16647	1901	Marindin	Tilting reflector
18273	1902	Barr & Stroud	Single telescope type
1462	1903	" "	Separating prisms, optical squares
4258	1903	Forbes	Stereoscopic
12735	1904	Dennis Taylor	Single telescope: swinging deflecting prism
28866	1904	König	Adjusters
11726	1905	Carl Zeiss	Separating prisms
10039	1906	Barr & Stroud	Deflecting rotating prisms
28728	1906	" "	Self-contained adjusters
12177	1907	Carl Zeiss	Periscopic
15200	1907	Dennis Taylor	Separating prism
16467	1907	Carl Zeiss	Stereoscopic
26546	1907	" "	Adjuster
830	1908	" "	Adjuster
1290	1908	" "	Eppenstein—unequal magnification fields
3526	1908	" "	Stereoscopic image marks
9906	1908	Barr & Stroud	Separating prism
13813	1908	" "	Handles and working heads
21854	1908	C. P. Goerz	Separating prism
22102	1908	Christie	Adjuster
22363	1908	Goerz	Separating adjuster prism
23173	1908	"	Separating prism
7785	1909	Barr & Stroud	Frames
7786	1909	" "	Scales and scale-erecting prisms
16847	1909	C. P. Goerz	Unequally divided fields
18611	1909	"	Separating prism
21870	1909	Carl Zeiss	Adjuster
30152	1909	Barr & Stroud	Separating prism
6082	1910	Dennis Taylor	Binocular coincidence
7392	1910	" "	Mirrors for optical squares
24714	1910	C. P. Goerz	Vertical and horizontal eyepieces
28022	1910	Carl Zeiss	Folding
513	1911	" "	Upper invert. double separating line
24821	1911	C. P. Goerz	Separating prism

[1] It will be noticed that this list is practically limited to patents taken out by the four chief manufacturing firms.

LIST OF THE MORE IMPORTANT BRITISH PATENTS FOR RANGE-FINDERS—*continued.*

Specification No.	Year of Application.	Name.	Subject-matter.
25122	1911	Carl Zeiss	Complementary inverted fields
25269	1911	C. P. Goerz	Separating prism
25368	1911	,,	Separating prism
12962	1912	Carl Zeiss	Halving adjustment
14041	1912	,, ,,	Adjuster
14145	1912	C. P. Goerz	Adjuster
16837	1912	Carl Zeiss	Adjuster
21027	1912	,, ,,	Periscopic
23696	1912	,, ,,	Periscopic
26233	1912	Barr & Stroud	Variable base range-finder
27483	1912	Carl Zeiss	Stereoscopic: with separating prism
30090	1912	Dennis Taylor	Different vertical and horizontal magnifications
103	1913	Carl Zeiss	Adjuster
1535	1913	,, ,,	Different vertical and horizontal magnifications
4879	1913	,, ,,	Binocular coincidence: different magnifications
8098	1913	,, ,,	Adjuster
10966	1913	,, ,,	Stereoscopic: with double images in each field
14183	1913	,, ,,	Separating prisms
15815	1913	,, ,,	Coincidence and stereoscopic: double deflecting devices
27217	1913	Barr & Stroud	Scale-indicating gear
8312	1914	Carl Zeiss	Grouped to give single reading
11463	1914	,, ,,	Optical squares
14701	1914	,, ,,	Optical squares
15140	1914	Barr & Stroud	Exhausting adjuster
15294	1914	,, ,,	Scale-conversion gear
20263	1914	Carl Zeiss	Adjuster
21280	1914	,, ,,	Addition to 15815/13
522	1915	,, ,,	Adjuster
2114	1915	Barr & Stroud	Exhausting range-finder
3544	1915	Carl Zeiss	Variable resistance scale indicator
1179	1915	,, ,,	Addition to 15815/13
4567	1915	,, ,,	Setting gun-sight
5426	1915	,, ,,	Adjuster
7679	1915	,, ,,	Friction elevating gear
14508	1915	,, ,,	Addition to 20263/14
125164 [1]	1916	Rottenburg & Willans	Height-finder
125587	1916	Barr & Stroud	Height-finder
125725	1918	Taylor	Adjusting end squares
127885	1917	Barr & Stroud	Separating-prism systems
127907	1917	,, ,,	Height-finder
129043	1917	Taylor	Single telescope: swinging-prism type
129044	1917	,,	Halving systems
129345	1917	Barr & Stroud	Height-finder: conversion gear
129881	1918	,, ,,	Height-finder: travelling deflecting prism
131610	1917	,, ,,	Height-finder: travelling deflecting prism
131611	1917	,, ,,	Height-finder: rotating deflecting prism
133974	1917	,, ,,	Scale-conversion gear
135223	1916	,, ,,	Multi-magnifying systems
139224	1916	,, ,,	Height-finder
145094	1916	,, ,,	Unit-magnification system
147106	1920	Carl Zeiss	Periscopic range-finder
149326	1920	,, ,,	Adjuster: zero base
165461	1918	Barr & Stroud	Wandering-picture graticule
177598	1920	,, ,,	Mounting large range-finders of the coast-defence type
188930	1921	,, ,,	Periscopic range-finders

[1] From 1916 the specification numbers run continuously, commencing with No. 100001. Year numbers are given provisionally.

F. J. C.

ROUSSEAU DIAGRAM: a device for obtaining the average candle-power of a source from its polar diagram. See "Photometry and Illumination," § (42).

ROWLAND'S GRATING. See "Diffraction Gratings, Theory of," § (11).

RUBENS AND WOOD, their improvement in the method of isolating very long waves, based on the selective refraction and selective absorption of quartz. See "Wave-lengths, The Measurement of," § (7).

RULING ON GLASS. See "Graticules."

RUMFORD PHOTOMETER. See "Photometry and Illumination," § (25).

RUSSELL ANGLES: a system of angles by means of which the calculation of average candle-power of a light source may be simplified. See "Photometry and Illumination," § (42).

RYDBERG, his relations between the wave-numbers of lines in the spectra of elements, and recognition of the existence of three chief types of series, known as the Principal, Sharp, and Diffuse series. See "Spectroscopy, Modern," § (10) (ii.).

— S —

SACCHARIMETER: a term applied to polarimeters, used to measure the strength of sugar solutions, in which the quartz wedge compensation system is employed. See "Saccharimetry," § (2).

SACCHARIMETER SCALES. See "Saccharimetry," § (4).

SACCHARIMETERS, conditions governing the accuracy and sensitiveness of. See "Saccharimetry," § (7).

Mechanical construction and design of. See *ibid.* § (6).

SACCHARIMETRY

§ (1) INTRODUCTORY.—The optical rotation of a sugar solution can be measured by means of the polarimeter,[1] and if the rotation of a *pure* sugar solution of the same concentration be known, the percentage purity of the first sample can then be calculated. Owing to the rotatory dispersion (variation of rotatory power with wave-length) of all active substances, a monochromatic light source is necessary in all polarimetric measurements. The half-shadow angle of the polariser must be small in order to obtain accurate results. But even with a half-shadow angle of 6°, only a small fraction of the incident light enters the eye of the observer at the matching point. The usual monochromatic light source (sodium light) was therefore too feeble for accurate measurements even when the optically active substance was highly transparent, and this is by no means the case with the average raw sugar solution.

Biot had observed that the rotatory dispersion of quartz was approximately the same as that of a sugar solution, and in response to a prize of 2000 francs offered by the "Société d'Encouragement" for a means of determining the strength of sugar solutions correct to 2 per cent, Soleil[2] ingeniously evolved the quartz compensator system, in which he utilised the rotatory power of quartz, to neutralise that of the sugar solution. With this arrangement, the polariser and analyser are *fixed and crossed*, and the rotation of a sugar solution is given by the distance a quartz wedge has to be moved in order to neutralise the rotation of the solution. Since the rotatory dispersions of quartz and of sugar solutions are approximately equal, this system permits the use of white light with its relatively great intensity and convenience.

§ (2) SACCHARIMETERS.—The term "saccharimeter" is now generally applied only to polarimeters in which this quartz wedge compensation system is used.

The optical arrangement in a standard polarimeter is shown in *Fig.* 1. The polarising prisms CD form the Lippich[3] polariser. The wedges K and L are made of left-rotating quartz and have their bases on opposite sides, so that a beam of light will pass through undeviated. The analysor M is "crossed" with the polariser CD before the quartz wedges are mounted. A compensating disc J of

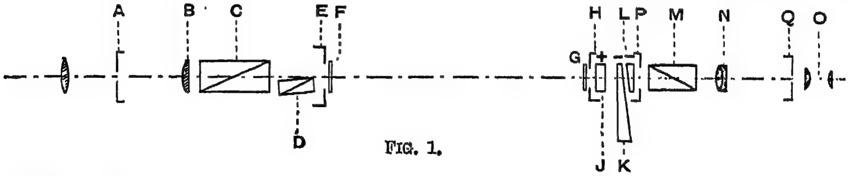


FIG. 1.

[1] See "Polarimetry."

[2] Soleil, *Compt. Rend.*, 1845, xx. 1805; 1845, xxi. 426; 1847, xxiv. 973; *Mém. Inst.*, 1845, xiii. 214.

[3] See "Polarimetry," § (6).

right rotating quartz is introduced. Its thickness is such that it equals that of the left rotating wedges K and L when the movable wedge K has been moved approximately one-fifth of its total path (as measured from the position when the *thinner* end of K is nearer the centre). The objective N is achromatic and the eyepiece O is focussed on the sharp dividing line of the prism D.

The light source is focussed on the diaphragm A of the polariser by a lens (shown in the diagram). A can therefore be regarded as

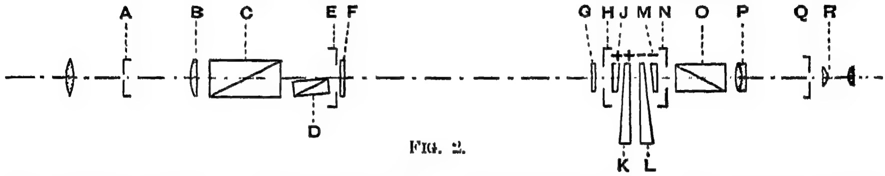


FIG. 2.

the virtual light source. An image of this is brought to a focus, in turn, at the objective N of the observing telescope, by means of the condenser lens B. Cover glasses F and G serve to protect the polarising and analysing systems from dust and damp.

When an observation tube containing a liquid is placed in the path between F and G, a halo is usually seen surrounding the usual field of view. This is caused by reflections at the inner walls of the tube; if a diaphragm is placed at Q (the focal plane of the eyepiece) so that its aperture is equal to, or only slightly greater than the size of the image of E formed by the lens N at Q, then this effect is avoided. The apertures H and P serve to cut off scattered light from outside and multiple back reflections.

In the double wedge instrument shown in *Fig.* 2 the optical arrangement is the same, only that now the quartz compensating disc (in the case of the single wedge instrument) is replaced by two wedges J and K of opposite rotation to the other pair L and M. The long

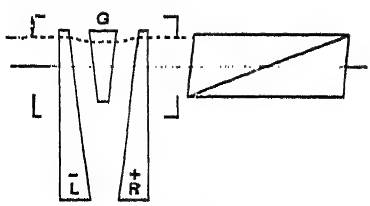


FIG. 3.

wedges K and L are movable, whilst the short wedges J and M are fixed. The special advantages of having two movable wedges of opposite rotations will be discussed later. Martens[1] suggested that the same result may be obtained as shown in *Fig.* 3. The one long wedge is right rotating while the other is left rotating. The deviation of the beam is compensated by means of the glass wedge G. Although this method means a considerable saving owing to the scarcity of optically pure quartz, it has not come into general use. There are possibly two reasons for this—first, the double refraction almost always present in a thick glass wedge with the decrease in sensitiveness that follows; and, second, the fact that the dispersion of glass being different from that of quartz a coloured fringe would be seen at each side of the field of view, since the light source is approximately white.

The standard instruments of Frič, Schmidt and Haensch, Goerz, Peters, Bausch and Lomb, and Hilger, are all optically identical with either *Fig.* 1 or *Fig.* 2. The saccharimeter of Pellin embodies a Laurent polariser, while in that of Bellingham and Stanley the modified Jellet prism described on p. 483 is employed. When a separate light source is used, another

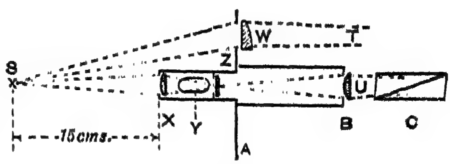


FIG. 4.

short focussed convex lens is mounted in front of the diaphragm A (*Figs.* 1 and 2), so that it forms an image of the source at A, when the latter is placed a particular distance (15 cm. in the Schmidt and Haensch instruments) in front of this auxiliary lens X (in *Fig.* 4). In order to secure a uniformly illuminated field of view a diffusing screen is placed at Z, and the image of S can be accurately focussed on it by looking sideways through the elongated observation hole Y at the side of the tube carrying the lens X. The diaphragm at A is extended outwards, as shown, to screen off the extraneous light from the observer. A deviation prism W is mounted in an aperture in this extended diaphragm so that a beam of light is obtained to illuminate the scales of the instrument. When electrical illumination is used, the large diaphragm at A and the deviating prism can be dispensed with, the illuminating beam T being obtained by placing a small mirror at the requisite angle above the source.

[1] Martens, *Zeitschr. für Instkde.*, 1900, xx. 82.

It must be noted that, as in the polarimeter, the beam emerging from the condenser B is not strictly parallel, but is made to converge on the objective of the observing telescope. This secures that the maximum permissible amount of light enters the eye of the observer, simultaneously with uniform illumination of the field of view.

§ (3) BATES'S SACCHARIMETER.—It is not possible to vary the half shadow angle of an ordinary saccharimeter, because it would entail a removal of the quartz wedges, etc., in order to set the analysing nicol at the new position of matching. If the polariser and analyser are not accurately crossed, a match cannot be made at any position of the wedge when the quartz compensating system is inserted, because, approximately white light being employed, the rotatory dispersion of quartz will cause considerable colour differences in the two halves of the field of view.

Bates [1] has constructed a saccharimeter in which the adjustment of the analyser is automatically made as the half shadow angle is varied by rotating the whole Lippich prism of the polariser.

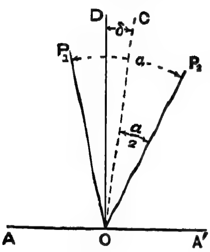


FIG. 5.

Let OP_1 in *Fig.* 5 represent the plane of polarisation of the light emerging from the whole Lippich prism, OP_2 the plane of polarisation of the half Lippich, and AOA′ that of the analysing nicol. OD is drawn perpendicular to AOA′, and OC bisects the angle P_1OP_2. Let δ be the angle that the normal to the plane of the analyser makes with OC, the bisector of the half shadow angle.

$$\angle P_1OA=\frac{\pi}{2}-\frac{\alpha}{2}+\delta,\quad \angle P_2OA'=\frac{\pi}{2}-\frac{\alpha}{2}-\delta.$$

If I is the intrinsic intensity of illumination of the light emerging from the whole Lippich, its intensity after transmission through the analyser will be given by

$$I\cos^2 P_1OA=I\sin^2\left(\frac{\alpha}{2}-\delta\right).$$

The intensity of the light leaving the half Lippich will be appreciably less. A portion is lost by reflection at the two ends of the prism, by absorption in the spar and by reflection at the balsam film; suppose this fraction to be $1-\kappa$; the fraction transmitted will be κ. Another portion is cut off because the half Lippich is not parallel to the whole Lippich, and the fraction transmitted will be given by $\cos^2\alpha$ according to the law of Malus.

The resultant intensity of the light from the half Lippich will therefore be $I\times\kappa\cos^2\alpha$, and this after transmission through the analyser becomes

$$I\times\kappa\cos^2\alpha\cos^2 P_2OA'=I\kappa\cos^2\alpha\sin^2\left(\frac{\alpha}{2}+\delta\right).$$

When a match of the fields is obtained, then

$$I\sin^2\left(\frac{\alpha}{2}-\delta\right)=I\kappa\cos^2\alpha\sin^2\left(\frac{\alpha}{2}+\delta\right),$$

$$\therefore \sin\frac{\alpha}{2}\cos\delta-\cos\frac{\alpha}{2}\sin\delta=\sqrt{\kappa}\cos\alpha\sin\frac{\alpha}{2}\cos\delta + \sqrt{\kappa}\cos\alpha\cos\frac{\alpha}{2}\sin\delta;$$

$$\sin\frac{\alpha}{2}\cos\delta(1-\sqrt{\kappa}\cos\alpha) =\cos\alpha\sin\delta(1+\sqrt{\kappa}\cos\alpha),$$

$$\therefore \tan\delta=\frac{1-\sqrt{\kappa}\cos\alpha}{1+\sqrt{\kappa}\cos\alpha}\tan\frac{\alpha}{2}.$$

If the loss of light by absorption and reflection at the half Lippich is neglected, then $\kappa=1$ and $\tan\delta=\tan^3\alpha/2$ as was shown by Bates in his first paper. But this factor κ is important, as was realised by Schönrock [2] and Wright [3] independently. The latter assumed that the loss of light in the half Lippich was approximately 10 per cent, the value of the constant κ in the above equation being therefore ·9. The following table, due to Wright, shows the difference introduced by considering this factor.

Half Shadow Angle in Degrees.	Value of δ from Bates's Formula.	Value of δ from Wright (κ =·9).	Half Shadow Angle in Degrees.	Value of δ from Bates's Formula.	Value of δ from Wright (κ =·9).
0	0	0	8°	1′	7′
2°	0	2′	10°	2′	10′
4°	0	3′	12°	4′	14′
6°	1′	4′	14°	6′	18′

Schönrock calculated the loss of light from the well-known equation of Fresnel. For perpendicular incidence the percentage loss is given by $(n-1)^2/(n+1)^2\times 100$. Taking n as 1·486 (refractive index of Iceland spar) the percentage loss is 7·5. A further allowance of ·5 per cent is arbitrarily made for absorption inside the small Lippich, thus making the constant κ in the above equation ·92. The effect of considering the absorption is, as Bates [4] pointed out, to make the variation of δ with α more approximately linear. If the analysing nicol is therefore made to rotate with an

[1] Bates, *Bull. Bur. Stand.*, 1907, iv. 461.

[2] Schönrock, *Zeitsch. Ver. Deut. Zuckerind.*, 1908, lviii. 111.

[3] Wright, *Amer. Journ. Science*, 1908, xxvi. 391.

[4] Bates, *Bull. Bur. Stand.*, 1908, v. 193.

angular velocity slightly greater than half of that of the whole Lippich, the variation of the zero point of the analyser can in Bates's construction be reduced to $\pm\cdot14°$ (or $\pm\cdot05°$ S.) from the true matching position as given above, even when the half-shadow angle is altered from 3° to 15°. But it appears that, where the highest accuracy is required in measuring a rotation, a small allowance must be made for the change of the zero position with the particular half-shadow angle employed.

§ (4) SACCHARIMETER SCALES.—The amount of the rotation of a substance is proportional to the distance that the movable wedge has to be displaced in order to neutralise the rotation. A scale is therefore fixed to the movable wedge, and a vernier mounted on the fixed one as shown in *Fig.* 6. The scale and vernier (S and V of *Fig.* 6) are illuminated by the beam from the deviating prism (T in *Fig.* 4). This falls on a grey glass screen W (*Fig.* 6)

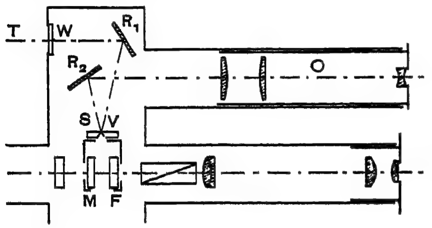


FIG. 6.

and is reflected by the mirror R_1 on to the scale and vernier. The reflected image of the latter in the mirror R_2 is observed by the telescope O.

It will be realised that the scales of all saccharimeters are arbitrary, depending on the angle of the wedge and on the length of a scale division, the rotation in a particular position being $nl \tan \theta(\kappa)$ where n denotes the number of divisions the wedge has been moved from its zero position, l the length in millimetres of a single scale division, θ the angle of wedge, and κ the rotatory power of quartz per millimetre for a particular wave-length.

The wedge angle and length of scale division adopted by Soleil were such that the rotation of the plane of polarisation of sodium light was 21·67° when the wedge was moved 100 divisions. This was supposed to be the amount by which the plane of polarisation of sodium light was rotated when passed through a quartz plate exactly 1 mm. thick, the faces of the plate being cut normally to the optic axis.

The value of the rotation given above is that of Broch (1846). Later determinations have shown that the specific rotation of quartz is appreciably greater than the above value; nevertheless, the 100° point on the scale has been standardised for the rotation as given by Broch, so that all French saccharimeters give almost identical readings—which is by no means the case with the other continental saccharimeters.

§ (5) ROTATION CONSTANTS OF SUGAR.—The specification of the normal sucrose solution which, when placed in a 200 mm. tube, would have a rotation of 21·67° for the sodium D lines at 20° C., has been the subject of a large number of determinations. The normal weight is defined as the number of grams of pure sucrose that is contained in 100 c.c. of the normal solution at 20° C., and values ranging from 16·0 grm. (Dubrunfaut) to 16·51 grm. (Clerget) have been given. In 1875, the value of 16·19 grm. (Girad, and de Luynes) was adopted as the official normal weight. The International Congress of Applied Chemistry (Paris, 1896) suggested that a redetermination should be made. The French Finance Ministry, acting on the report of Mascart and Benard,[1] adopted the value 16·29 grm., which remains up to the present the official French normal weight.

Ventzke[2] in 1842 proposed that the rotation of a standard sugar solution should be used to determine the 100° point on a saccharimeter scale. Originally, a 25 per cent solution of cane sugar was suggested as the normal solution; he afterwards proposed that the specific gravity of such a solution could be used for determining the concentration (that the normal solution should have a specific gravity of 1·1 at 17·5° C.). It was found impossible to make determinations of the specific gravity with sufficient accuracy, so that the normal solution was defined as one containing 26·048 grm. of sucrose in 100 c.c. of the solution at 17·5° C.[3]

In 1855 the Mohr c.c. came into general use (1 Mohr c.c. = volume occupied by 1 grm. of water weighed in air with brass weights at 17·5° C., 100 Mohr c.c. = 100·234 metric c.c.), and, until 1900, saccharimeters were standardised with a normal solution based on the Mohr c.c. whilst the normal weight remained unaltered. Thus the scale usually recognised as that of Ventzke has a normal weight 1·00234 times greater than that proposed by him.

Owing to the confusion and errors resulting from the two standards of volume, the International Commission for uniform methods of sugar analysis[4] recommended a new definition based upon the metric c.c. and a standard temperature of 20° C., as 17·5° C. is generally below the average temperature of a sugar laboratory. After correcting for the change

[1] Mascart et Bénard, *Ann. Chim. Physique*, 1899, (7), xvii. 127.
[2] Ventzke, *Erdman's Journ. für praktische Chemie*, 1842, xxv. 65.
[3] *Ibid.*, 1843, xxviii. 111.
[4] *Zeitschr. Ver. Deut. Zuckerind.*, 1900 (I.) l. 357.

in the specific rotation of the sugar solution with temperature, and allowing for the temperature coefficient of the saccharimeter and tube, a normal weight of 26·01 grm. is obtained for an instrument standardised at 17·5° C. with the Mohr flask.

The International Commission decided to make the new normal weight exactly 26·000 grm. The following definition was formulated, and has since been accepted in all countries with the exception of France and the French colonies.

"One dissolves (for instruments arranged for the German normal weight) 26·00 grm. of pure sugar in a 100 metric cubic centimeter flask, weighing to be made in air with brass weights, and 'polarises' the solution in a room the temperature of which is 20° C. Under these conditions the instrument must indicate exactly 100·00.

"The temperature of all sugar solutions to be tested is always to be kept at 20° C., while they are being prepared and while they are being 'polarised.'"

By "polarising" is understood in this connection the measurement of the number of scale divisions that the wedge has to be moved in order to secure a match of the fields of the Lippich polariser when a column of the normal solution 200 mm. long is placed in the path, the point of reference being the reading on the scale when the fields are of equal intensity, *without* any active solution in the path.

Owing to difficulty of making a solution of pure sucrose of the required concentration, a secondary standard has to be made for both the manufacturer in the primary graduation of the instrument, and the user of the instrument for checking purposes.

The subsidiary standard is a quartz plate which when measured with a saccharimeter gives the same reading as that of a normal solution with the same instrument. The angular rotation of such a plate at 20° C. for sodium light was determined by Schönrock[1] in co-operation with Herzfeld,[2] the value given by him being 34·657°. Practically all saccharimeters employing the scale of the International Sugar Commission have, until recently (1917), been standardised on this basis. The above figure is commonly called the "conversion factor," and its accurate determination is of great importance in saccharimetry.

It had been noted at the Bureau of Standards that all normal sugar solutions, however carefully prepared, gave a reading below 100·00° when polarised in saccharimeters standardised on the basis of the conversion factor given by Herzfeld and Schönrock.

[1] Schönrock, *Zeitschr. Ver. Deut. Zuckerind.* (Techn. Teil), 1904, liv. 521, abst. in *Ann. d. Phys.*, 1904 (4), xiv. 406.

[2] Herzfeld, *Zeitschr. Ver. Deut. Zuckerind.* (Techn. Teil), 1900, l. 826.

A preliminary series of measurements were undertaken by Bates and Jackson[3] in 1912, when they found that the normal solution read 99·90° S. This was subsequently verified by Walker,[4] who prepared sucrose by the alcohol method and obtained 99·88° S. This led Bates and Jackson[5] to undertake an exhaustive investigation of the whole problem.

The absolute rotation of a normal solution of sucrose in a 200 mm. tube was measured by means of the spectro-polarimeter, using the green mercury radiation. This was found to be 40·763°. The absolute rotation of a quartz plate that gave the same reading on a saccharimeter as the normal solution was 40·690°, the difference being due to the slightly different rotatory dispersions of the two substances.

The ratio of the angular rotations of a quartz plate for sodium light (λ = 5892·5 Å.) and the green mercury radiation (5461 Å.) had been previously measured by Bates[6]; this constant was re-determined and the mean value found to be

$$\frac{\phi^{20^\circ \text{C.}}_{\lambda=5892\cdot5\,\text{Å}}}{\phi^{20^\circ \text{C.}}_{\lambda=5461\,\text{Å}}} = \cdot85085.$$

This ratio for the normal sugar solution was also measured and found to be ·84922.

The value of the conversion factor is therefore 40·690 × ·85085 = 34·620° for sodium light (λ = 5892·5 Å.) as against the value 34·657° from the Herzfeld-Schönrock determination.

Herzfeld[7] has criticised this result on the ground that Bates and Jackson have omitted to guard against the presence of micro-organisms in their standard solutions, and that the solutions from which the pure sucrose was crystallised were not kept slightly alkaline during evaporation to avoid slight hydrolysis. As Bates and Jackson have pointed out in their paper, the value for the specific rotation of sugar for the normal solution obtained from the Herzfeld-Schönrock conversion factor is very much higher than the average value determined from the formulae of Tollens and of Nasini and Villavecchia as combined by Landolt.[8] According to the latter, $[\alpha]_D^{20^\circ \text{C.}} = 66\cdot502$, whilst the value given by Schönrock[9] is 66·627. The value obtained from the redeterminations of Bates and Jackson is 66·529, which serves to corroborate the

[3] Bates and Jackson, *Eighth International Congress of App. Chem.*, 1912, xxv. 517.

[4] Walker, *Sugar*, 1915, xvii. No. 2, 47.

[5] Bates and Jackson, *Bull. Bur. Sta.*, 1916, xiii. 67.

[6] Bates, *ibid.*, 1906, ii. 247.

[7] Herzfeld, *Zeitschr. Ver. Deut. Zuckerind.*, 1917, lxviii. 407.

[8] Landolt, *Das optische Drehungsvermögen* (2nd ed.), 1898, p. 420.

[9] Schönrock, *Zeits. Ver. Deut. Zuckerind.*, 1904, liv. 553.

value that they have obtained for the conversion factor.[1]

This modification of the Herzfeld-Schönrock values for the scale of the International Commission has been accepted at the National Physical Laboratory (England) and at the Bureau of Standards (America). Apparently pending international confirmation it has not yet been accepted by other bodies.

Sidersky and Pellet (1896) suggested that 20 grm. should be used as the normal weight instead of the 16 grm. (French) and the 26 grm. (German). The proposal was rejected by the International Commission[2] in 1906. A strong movement for its adoption was started in America[3] in 1918. The reasons in favour of, and against, this scale are considered by Saillard.[4]

Up to the present, however, no reliable data of sufficient accuracy are available to justify its adoption for high-precision saccharimeters.

§ (6) Construction and Design of Saccharimeters.—The polariser and analyser (with quartz compensating arrangement) are usually mounted on rigid trestle supports fastened to a heavy base. The polariser is similar to the types used on polarimeters (*q.v.*) with fixed half-shadow angle (with the exception of the Bates-Frič instrument). The movable wedge of the compensating system is moved along close-fitting ways by means of a rack-and-pinion movement. The scale is mounted on the carriage holding the movable wedge. While the other wedge is fixed, in the sense that it is not moved when an observation is being made, it can be moved laterally by means of a screw and key when necessary in order to arrange that at zero point of matching, the lengths of the movable wedge on each side of the optic axis of the instrument should be in the proportion of 30 : 115. This is needed so that the full length of the wedge and scale can be utilised. The vernier scale can be moved independently in a similar manner to correct for the zero error of the scale. The method of illuminating and observing the scale is indicated in *Fig.* 6. When two sets of wedges are employed as in *Fig.* 2, two separate observing telescopes are used in the Frič instruments, while in those of Goerz, and Schmidt and Haensch, the two sets of scale and vernier are brought into the field of view of the one observing telescope.

The range covered by single wedge (movable) saccharimeters generally extends from − 30° S. to + 110° S. or 115° S., the verniers reading directly to ·1° S. (in the case of the Hilger saccharimeter the vernier reads directly to ·05° S.). Generally it is possible to estimate to within a half of the vernier division. The scales and vernier are made of nickeline having a coefficient of expansion of ·000018, or of glass (·000008). A sensitive thermometer is often mounted in the compensator head, so that the temperature of the wedges and scales can be accurately known.

The movable wedge of the Bellingham and Stanley saccharimeter is only one quarter of the length of the usual wedge, and therefore the scale is correspondingly shorter. The latter is engraved by means of a photoceramic process and is observed by transmitted light, by means of a low-power microscope. The vernier is similarly engraved on a glass plate and is mounted in the focal plane of the microscope ocular.

As the above-named process admits of only one definite size for the scale and vernier, the variations in the angle of the wedge are compensated for by a slow motion to rotate the wedge in the plane perpendicular to the optic axis of the instrument, since it is the angle of the wedge in the direction of its motion that is the effective factor in determining the length of the scale.

The external features and other details of mechanical construction are best obtained from the makers named.

§ (7) Conditions governing the Accuracy and Sensitiveness of Saccharimeters.—The optical purity of the quartz in the compensating system to a great measure controls the accuracy and sensitiveness of a saccharimeter. Apart from regular twin crystals, which can often be detected by an examination of the crystal faces,[5] it is generally found that a parallel plate of quartz cut perpendicular to the optic axis contains a number of spike-shaped crystals of opposite rotation irregularly distributed, usually around the outside of the plate and pointing towards the centre.

These can be conveniently detected by the method of Buisson,[6] where interference bands are produced by the quartz plate in a parallel beam of polarised light. When green mercury light is employed the interpenetrating crystals show up as sharp discontinuities in the otherwise uniform interference bands.

If a quartz plate is examined in a powerful polarised beam of white light (say from a point-o-lite lamp) and the rotation of the plate be compensated for by adjustment of a pair of quartz wedges placed directly in front of the analysing nicol (method due to Twyman) a series of bands are seen of a hexagonal shape parallel to the natural edges of the plate.

[1] An independent determination of the 100° point of the Herzfeld-Schönrock scale has been made recently by Stanek. His results confirm the values obtained by Bates and Jackson. *Listy cukrovarnicke*, 1920-21, No. 45; *Zeitsch. Zuckerind. cechoslovak Republik*, 1921, xlv. 417, 425.

[2] *Inter. Sugar Journ.*, 1907, lx. 5.

[3] Cf. Browne, *Louisiana Planter*, Dec. 7, 1918; Jan. 5, 1919.

[4] Saillard, *Journ. des Fabricants de Sucre*, 1919, lx., No. 13.

[5] Cf. Lewis, *Crystallography*, p. 519.

[6] Buisson, *Journ. d. Phys.*, 5 série, 1919, ix. 25.

These apparently indicate that the rotatory power of quartz is to some extent influenced by conditions at the time of its growth. As these bands are invariably present, even when the above-mentioned interpenetrating crystals are absent, it is important to cut the wedges in a direction *along* the band.

It is owing to the lack of homogeneity of quartz that the double wedge system (*Fig.* 2) is of considerable advantage. Not only can the scale of working wedge be calibrated with reference to that of the "control" wedge at all points along its path, but also by placing the latter at different points, the rotation of a sugar solution can be measured with different portions of the working wedge, thereby eliminating errors due to local imperfections in the wedge.

These local errors in saccharimeter wedges can also be determined by means of the "control tube" of Schmidt and Haensch.[1] This consists of a telescopic observation tube of metal which can be adjusted by means of a rack-and-pinion movement, to give a column of solution of any length between 225 and 410 mm. By means of a vernier the exact length of a column of solution within this range can be read upon a scale, to within 0·1 mm. If an approximate half-normal solution is used, the 100° S. point on the saccharimeter will be given by a column about 400 mm. long. If the length of column required to neutralise the wedge at each scale division is determined, a calibration curve is obtained that allows for the faults of the wedge and also for any inequalities in the scale divisions. It is of course important that the scale of the telescopic tube be accurately calibrated independently. Browne[2] has given the curves of errors of a Schmidt-Haensch and a Bates-Frič saccharimeter taken annually over a period of seven years. It is interesting to note that while local imperfections persist, there seems to be a small progressive decrease, from year to year, in the average error of the scale.

If the optic axes of the quartz wedges and compensating disc do not accurately coincide with the direction of the optical axis of the instrument, the incident plane-polarised beam becomes elliptically polarised after transmission through the quartz, to an extent dependent on the amount of "axis error" and the thickness of quartz traversed. The methods of Gumlich[3] and Walker[4] seem only satisfactory for comparatively thick plates. Schönrock[5] has investigated the interference bands (in convergent polarised light) of a plate of quartz the optic axis of which makes a small angle with the normal to the plate. In conjunction with Brodhun[6] he has evolved a method of accurately determining the axis error of a parallel plate of quartz. The quartz plate is mounted on the table of a small spectrometer, and is mounted at an angle of about 40° to the optical axis of the collimator and telescope, which are in alignment. Nicol prisms are placed in both the telescope and collimator, and on illuminating the slit with sodium light the dark interference bands are seen. The plate is rotated in its own plane, until the displacement of the bands in either direction is a maximum. The plate is turned (about a vertical axis) until any one band is central with the cross wires in the telescope (or to the slit image). The plate is now rotated through 180° in its own plane, the plate being simultaneously turned (by rotating the table) so as to keep the band in the same position. The axis error of the plate is given by one-third of the angle that the plate has to be rotated. By employing a slow-motion movement for turning the spectrometer table Brodhun and Schönrock find that axis error can be determined to within a few seconds of arc.

The object of correcting for axis error in the quartz is to avoid elliptic polarisation of the plane-polarised beam after transmission through the quartz. The same effect is caused by accidental double refraction at any point between the polariser and analyser. It is therefore equally important that all the wedges, etc., be mounted as free from strain as possible. This is particularly applicable to the quartz control plates used for standardising and checking the 100° point of the saccharimeter, as the temperature variations are more pronounced in this case, as Weichman[7] and others have observed. Williams[8] has devised a mount for the control plate, in which the latter is held freely in a guard ring of steel, ·0025 mm. thicker than the quartz plate.

The glass plates protecting the polariser analyser should be made of well-annealed optical glass, as should also the end plate or cover glasses of the observation tube. The latter are often made of plate glass, this being the origin of some of the perplexing differences in rotation values obtained for one and the same solution examined in different tubes.

Owing to the difference between the rotatory dispersions of sugar solution and quartz (which though small is not negligible), the colour difference between the two halves of the field at the matching point is quite appreciable when a white light source is used, and an accurate matching is very difficult. The difference in

[1] Schmidt und Haensch, *Zeitschr. für Instkde.*, 1884, iv. 169.
[2] Browne, *Journ. of Ind. and Eng. Chem.*, 1920, xii. 792.
[3] Gumlich, *Wiss. Abh. d. phys. techn. Reichsanstalt*, 1895, ii. 201.
[4] Walker, *Phil. Mag.*
[5] Schönrock, *Zeitsch. Instkde.*, 1902, xxii. 1.
[6] Brodhun und Schönrock, *Zeitsch. Instkde.*, 1902, xxii. 364.
[7] Weichman, *School of Mines Quarterly*, 1899, xx.
[8] Williams, *Inter. Sugar Journ.*, 1919, xxi. 330.

the values of the rotation of a normal solution when the saccharimeter is illuminated with green mercury light ($\lambda=5461$ Å.) and sodium ($\lambda=5892{\cdot}5$ Å.) is ·188° S. (Schönrock [1]) or ·185° S. (Bates and Jackson [2]), the value for the green mercury radiation being the higher.

Schönrock [3] therefore suggested that one of the sodium light filters of Lippich [4] should be used. The latter used a 10 cm. column of a 6 per cent solution of potassium bichromate, but Schönrock found that a 1·5 cm. layer was sufficient. The International Sugar Commission in 1912 adopted the resolution that whenever white light is used in saccharimetric observations it must be filtered through a solution of potassium dichromate of such concentration that the product length of column in cm. and percentage concentration =9.

The evaporation of the water and the instability of the colour necessitates that the solution be periodically renewed. To obviate this difficulty Adam Hilger, Ltd., have evolved an aniline dye filter that is more permanent and gives on a saccharimeter readings identical with those of the standard bichromate filter. *Fig.* 7 gives the percentage transmission curves

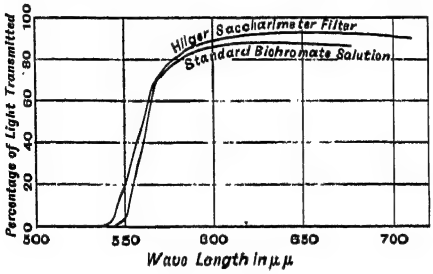


FIG. 7.

for the two filters, the data being obtained by means of a Hilger-Nutting spectro-photometer.

It will be noticed that these filters only exclude the lower portion of the spectrum, all wave-lengths above about 5700 Å. being transmitted freely. It is only to be expected therefore that different observers should obtain slightly different results owing to their varying spectral sensibility curves. An interesting observation has been made by Browne.[5] Three observers adjusted the length of control tube required to equalise the intensities in the fields of two saccharimeters the wedges of which were set at 100° S. Dividing the mean value of each observer (for each instrument) by the general average length of column, the following values, in terms of the saccharimeter scale, were obtained.

Observer.	Schmidt and Haensch Saccharimeter.	Bates-Frič Saccharimeter.	Average.
	° S.	° S.	° S.
A	100·020	99·971	99·9955
B	99·992	100·013	100·0025
C	99·987	100·017	100·0020
Average .	99·9997	100·0003	100·0000

It will be noticed that there is a difference of about ·04° S. between the results of A and C on the one instrument, while with the other instrument this difference has an approximately equal negative value. On examination it was found that the half-Lippich prism of the Bates-Frič saccharimeter is on the right, while in the Schmidt and Haensch instrument it is on the left. This indicates that the partial absorption of the light by the half-Lippich prism is an appreciable factor in the complex problem.[6]

§ (8) CORRECTIONS FOR TEMPERATURE.—As the temperature corrections in saccharimetry are unusually complicated, it is desirable to work at a temperature as near to 20° C. as possible.

The linear coefficients of expansion of quartz parallel and perpendicular to the optic axis are ·000007 and ·000013 respectively, so with a change of temperature the angle of the wedge is altered, the coefficient of the change being −·000006. The specific rotation of quartz has a temperature coefficient of ·000136 (Schönrock [7]), whilst that of the metal scale is about ·000018 (or ·000008 for glass).

The total temperature coefficient of a saccharimeter is ·000007 − ·000013 + ·000136 + ·000018 = ·000148 in the case of the metal scale, and ·000138 when the glass scale is used.

The true reading of the instrument at 20° C. is given by

$$R_{20} = R_t + R_t(t - 20°\text{ C.})\cdot 000148 \text{ (metal scale)},$$

or

$$R_{20} = R_t + R_t(t - 20°\text{ C.})\cdot 000138 \text{ (glass scale)},$$

where R_t is the reading of the saccharimeter at the temperature t.

The temperature coefficient of a quartz plate is ·000136 + ·000007 = ·000143. So that if P_t is the rotation in sugar degrees of the plate at $t°$ C. its rotation at 20° C. is $P_{20} = P_t + P_t(\cdot 000148 - \cdot 000143)(t - 20)$ for the

[1] Schönrock, *Zeitsch. Ver. Deut. Zuckerind.*, 1904, liv. 552.
[2] Bates and Jackson, *Bull. Bur. Sta.*, 1916, xiii. 123.
[3] Schönrock, *Zeitsch. Ver. Deut. Zuckerind.*, 1904, liv. 552.
[4] Lippich, *Zeitsch. Instkde.*, 1892, xii. 340.
[5] Browne, *Journ. of Ind. and Eng. Chem.*, 1920, xii. 792.
[6] According to a footnote to the above paper, Browne mentions that Horne has experienced a similar reversal of the "personal equation" by looking at the image of the field in a plane mirror, or by observing the field with the head bent downward.
[7] Schönrock, *Zeitschr. Ver. Deut. Zuckerind.*, 1904, liv. 521.

metal scale, and $P_{20} = P_t + P_t(\cdot000005)(20 - t)$ for the glass scale saccharimeter.

Schönrock[1] has determined the temperature coefficient of sucrose between 9° C. and 31° C., and in the neighbourhood of 20° C. it is ·000184, the coefficient of expansion of the solution is ·000285, and that of the glass observation tube ·000008, hence the total coefficient for the sucrose solution in the tube is ·000461. If we neglect the effect of temperature on the reading of the quartz plate, we can express the rotation of a sucrose solution at 20° C. (W_{20}) in terms of polarisation carried out at t° C., as

$$W_{20} = W_t + W_t\{\cdot000461 + \kappa\}\{t - 20\},$$

where κ is the temperature coefficient of the saccharimeter itself.

The temperature coefficients of other sugars are given by Browne.[2]

§ (9) Technical Methods of Saccharimetry.—The saccharimeter is most generally used in the analysis of the products of the cane- and beet-sugar industries. Owing to the similarity in their rotatory dispersions, nearly all the sugars can be quantitatively estimated by means of the saccharimeter, although in certain instances, *e.g.* commercial glucose, dextrin, etc., it is advisable to use a bichromate filter of double strength (*i.e.* percentage concentration × length in cm. = 18), as the rotatory dispersion of these substances is greater than with sucrose.

For full details of the chemistry and technique of sugar analysis, reference must be made to the standard text-books of Lippman,[3] Browne,[4] Weichman,[5] and others, and to the Circular (No. 44) of the Bureau of Standards. The following is the usual method of carrying out a direct polarisation, *i.e.* the determination of sucrose, in the absence of any other optically active substance.

Dissolve the normal weight (26 grm.) of the substance (assumed to be chiefly sucrose) in a 100 c.c. flask containing some water. In general the solution will be dark, and a clarifying agent must be added, usually 1-3 c.c. of basic lead acetate solution followed by 1-2 c.c. of alumina cream. (The amount of clarifying agent should be as small as possible.) The volume is then completed to the 100 c.c. mark, any foam accumulating on the meniscus and preventing accurate reading being dispersed by blowing on it a little alcohol or ether from an atomiser.

The solution is shaken thoroughly and filtered, the first portion of the filtrate being rejected, the remainder being well shaken up again. A 200 mm. tube is well rinsed in the solution and filled. The reading observed on the saccharimeter (after correction for zero error of instrument) will give directly the percentage purity of the sample. If the whole operation has been carried out at a temperature other than 20° C. a correction of ·03° S. per 1° C. difference can be used for high grades of sugars.

§ (10) Invert or Double Polarisation Method.—The method of direct polarisation, as explained above, gives the percentage of sucrose only in the absence of other optically active substances. The double polarisation method depends upon the principle that when sucrose is hydrolysed the reaction is expressed by the following equation:

$$\underset{\text{Sucrose.}}{C_{12}H_{22}O_{11}} + \underset{\text{Water.}}{H_2O} = \underset{\text{Glucose.}}{C_6H_{12}O_6} + \underset{\text{Fructose.}}{C_6H_{12}O_6}.$$

This resulting mixture of glucose and fructose in equal quantities is termed Invert Sugar, which has a specific rotation $[\alpha]_D^{20} = -20$. It will be seen that one part of sucrose is converted to $(360/342) = 1{\cdot}05263$ parts of invert sugar.

If the specific rotation of sucrose be taken as 66·5, then the ratio of optical activity before and after inversion will be

$$\frac{66{\cdot}5}{1{\cdot}05263(-20)} = \frac{-66{\cdot}5}{21{\cdot}0526},$$

which is a *decrease* of 87·5526 in the specific rotation, and the decrease for a normal sucrose solution would be $(100 \times 87{\cdot}5526)/66{\cdot}5 = 131{\cdot}66$, so that the scale reading of a normal solution of pure sucrose would after inversion read −31·66° S.

It follows, therefore, that the decrease in the saccharimeter reading on inversion when divided by the factor 1·3166 gives the percentage of sugar originally present. This of course is only the case when no other optically active substance present is hydrolysed, and when the agent used for inversion does not influence the specific rotation of the other substances. The inversion method of determining sucrose was discovered by Clerget[6] in 1849, the inversion agent being a few c.c. of hydrochloric acid. It is necessary to warm the solution to a temperature of 68° C., and on cooling again allowance must be made for the dilution of the invert sugar by the acid. When the solution after inversion is too dark to be measured on the saccharimeter, it may be decolorised by means of filtration through animal charcoal, or by the addition of reducing agents such as zinc dust, etc., the destruction of colouring matter being due to the nascent hydrogen liberated by the action of the hydrochloric acid present on the zinc.

Unfortunately, the hydrochloric acid considerably influences the specific rotations of

[1] Schönrock, *Zeitschr. Ver. Deut. Zuckerind.*, 1903, liii. 650.
[2] Browne, *Handbook of Sugar Analysis*, 1912, p. 128.
[3] Lippman, *Chemie der Zuckerarten*, 1904.
[4] Browne, *Handbook of Sugar Analysis*, 1912.
[5] Weichman, *Sugar Analysis*, 3rd ed., 1914.
[6] Clerget, *Compt. Rend.*, 1849, xvi. 1000.

fructose, raffinose, amino compounds, etc., that may be present, and the Clerget method is modified by either equalising the conditions before and after inversion with hydrochloric acid, or to use other inverting agents that do not have the objections of hydrochloric acid. Numerous organic acids have been tried, but as a whole have not been satisfactory. Kjeldahl[1] in 1881 employed yeast as the inverting agent, the active part of which is the enzyme invertase, which can be prepared by the method given by Hudson.[2]

This invertase method is probably the best of the Clerget modifications, as it has no effect on the specific rotations of substances usually present with sucrose except in the special cases where raffinose or gentianose is present.

As the temperature of a solution of invert sugar is raised, the specific rotation of the fructose is gradually decreased, whilst that of the other constituent (glucose) is not affected. Those two substances being present in equal quantities, a temperature can be reached at which the rotatory power of the one substance will exactly neutralise that of the other. This temperature is usually taken to be 87° C., and forms the basis of the method of Chandler and Ricketts,[3] originally employed for estimating the amount of commercial glucose present in cane-sugar, molasses, and honey. The saccharimeter reading is obtained by placing the solution in a water-jacketed observation tube (as described in "Polarimeter," § (17)), and circulating water from a thermostat until the reading of a thermometer placed at the central opening is steady, at 87° C.

W. E. F.
F. T.

Saccharimetry, direct polarisation method of. See "Saccharimetry," § (9).
Invert or double polarisation method of. See *ibid.* § (10).
Technical methods of. See *ibid.* § (9).
Temperature corrections in. See *ibid.* § (8).

Saturation (of Colour in Light), the distinctness or vividness of hue. See "Spectrophotometry," § (2).

Saturation, Degree of, used in connection with colour as an indication of purity. See "Eye," § (8).

Savart's Polariscope. See "Polarised Light and its Applications," § (15) (iv.).

Scale: a term used in music to denote a series of notes proceeding up or down at certain specified intervals. See "Sound," § (3).

Scale-reading Projection for Sensitive Instruments. See "Projection Apparatus," § (17).

[1] Kjeldahl, *Compt. Rend. Labor. Carlsberg*, 1881, i. 192.
[2] Hudson, *Journ. Indus. Eng. Chem.*, 1910, ii. 143.
[3] Chandler and Ricketts, *Journ. Am. Chem. Soc.*, 1880, ii. 428.

SCATTERING OF LIGHT BY GASES, WITH SPECIAL REFERENCE TO THE BLUE SKY

§ (1) Light from the Sky.—The fact that the sky is luminous at all calls for explanation. Its light is evidently derived from the sun, as the sky becomes dark whenever the sun is far below the horizon. But why do we receive sunlight from directions away from the sun, and why is the light modified to a blue colour?

If the sky in a direction at right angles to the sun is examined with a Nicol prism, it will be found that the light is strongly polarised. The direction of vibration is at right angles to the line joining the sun to the point examined. The polarisation, though by no means complete, is very conspicuous, and could not be missed by the most casual observer.[4] Capacity to explain this effect must be a crucial test of the merits of any proposed theory of the blue sky.

All the explanations which have from time to time been given postulate some material substance which intercepts part of the incident sunlight and scatters it laterally towards the observer on the earth's surface.

This lateral scattering has always been connected with the earth's atmosphere. The blue colour seen against the dark background of a distant mountain is obviously to be attributed to the same general causes as the blue of the zenith; thus we cannot assign the latter to anything that occurs in outer space.

From many points of view fluorescence of air would seem to afford an explanation. The ultra-violet light of the sun might give rise to a lateral scattering of blue light of lower refrangibility than its own. This theory, however, is negatived by the fact that the Fraunhofer lines are present in the light of the blue sky, just as in the direct light of the sun. A fluorescence spectrum has on the other hand no connection with, or at least no detailed resemblance to, that of the exciting light. Moreover, this view leaves the polarisation unexplained. Another theory was attempted by Clausius.[5] He assumed that bubbles of water were present in the upper atmosphere, and he regarded the blue colour as an example of the colours of thin plates, produced by reflexion from the bubbles. If these are sufficiently thin the colour will be the blue of the first order. This theory, however, fails to account for the richness of

[4] The phenomenon may be used to bring the moon in its earlier phases into view at an earlier hour of the day than would otherwise be possible: the masking light of the sky being partially quenched with the Nicol.
[5] *Pogg. Ann.* lxxii., lxxvi., lxxxviii.

sky blue, nor does it meet the requirements as regards polarisation very satisfactorily. Moreover, the presence of bubbles of the required thickness is a purely *ad hoc* supposition.

The true direction in which to look for an explanation was pointed out by the experiments of Brücke.[1] He poured a small quantity of a solution of gum mastic in alcohol into an excess of water, which has the effect of producing a precipitate of the gum, in a fine state of division, and capable of remaining in suspension for a long period. If this liquid is illuminated by a horizontal beam of sunlight, the light scattered laterally will show a marked blue colour. When this is examined with a Nicol prism at right angles to the primary beam, it is found to be polarised almost completely, the vibrations of the laterally emitted, or *scattered*, light being executed in the vertical plane. If, keeping in the horizontal plane, we pass away from the rectangular direction, the percentage of polarisation becomes less, tending to vanish as the original direction of propagation is approached. The conditions of symmetry round the original beam allow us to anticipate this, since there is nothing to distinguish vertical and horizontal directions of vibration in the limiting case.

An alternative method of producing the effects was employed by Tyndall,[2] who took advantage of a peculiar property of certain organic vapours. Butyl nitrite vapour, for instance, when mixed with a little hydrochloric acid, gives a fine cloud as the result of chemical decomposition when illuminated by a powerful beam from the sun or from the electric arc, and this cloud, for a minute or two after its formation, shows the blue lateral emission to good advantage: though the present writer has never been able to get so good an approach to a pure blue sky as Tyndall's descriptions would lead one to expect. As the action proceeds the particles become larger, and the scattered light tends to whiteness. At the same time the polarisation becomes less marked. The condition for a good blue, and for nearly complete polarisation, is that the particles should be small.

§ (2) LORD RAYLEIGH'S THEORY.—Small compared to what? The answer was given by the third Lord Rayleigh.[3] They must be small compared with the wave-length of light, so that at any given moment nearly the same phase of the incident vibration prevails at all parts of the particle.

Rayleigh entered on a full theoretical discussion of the subject.[4] He showed that if the particles are regarded as small independent vibrators, with their phases at random, simple considerations of symmetry indicated that the polarisation must be as observed. The case was essentially different from that of polarisation by reflexion from a glass surface, in that the phases of the elementary vibrators in that case were not at random. It was also simply shown from the theory of dimensions that the intensity of the scattered light must be inversely proportional to the fourth power of the wave-length: thus the short waves have a great predominance, which accounts for the blue colour. From a consideration of the analogy of waves on water, it will readily be seen that a small obstacle is more effective in breaking up and scattering short waves than long ones. For a more detailed treatment of the subject, Rayleigh made use of the elastic solid theory of light. Although this theory is now obsolete, it has sufficient formal analogy with the electromagnetic theory to form a trustworthy guide. Later, the subject was rediscussed by him in terms of the electromagnetic theory.[5]

The blue colour of the sky, then, is to be attributed to the scattering of light by small particles. But of what nature are these particles? The earlier writers appealed to atmospheric dust, and Rayleigh in his earlier papers took the same view, favouring especially particles of common salt. Later,[6] he showed theoretically that the molecules of the air itself would account for the greater part of the effect. The calculation was based on the values obtained by Bouguer, who examined the transmission of sunlight to the atmosphere, for various altitudes of the sun. The light scattered laterally is missing from the transmitted beam, and hence a measurement of the absorption gives a means of estimating the amount of, or at least a superior limit to, the scattering. In the calculation referred to, it was connected with the refractive index of air, and with the number of molecules in a cubic centimetre.

Later and much more accurate results for atmospheric transmission have been obtained by Abbot working on Mt. Wilson, and knowledge of the number of molecules in a c.c. is now much more definite. Using these improved data, Schuster[7] shows that the formula gives results within 1 or 2 per cent of those observed.

It is to be remarked that this accurate agreement is only got when the observations are taken at a high altitude. The lower part

[1] *Pogg. Ann.* lxxxviii. 303.
[2] *Phil. Mag.*, Series 4, xxxvii. 385; *Phil. Trans.*, 1870.
[3] *Phil. Mag.*, 1871, xli. 107-120, 274-279; *Collected Works*, i. 87.
[4] *Phil. Mag.*, 1871, xli. 447-454; *Collected Works*, i. 104.
[5] *Phil. Mag.*, 1881, xii. 81; *Collected Works*, i. 518.
[6] *Phil. Mag.*, 1899, xlvii. 375; *Collected Works*, iv. 397.
[7] *Theory of Optics*, 2nd ed. p. 329.

of the atmosphere—say the lowest mile—contains dust which adds considerably to the scattering due to the molecules themselves.

§ (3) Dust-free Air: Experimental Verification.—Thus far the investigation was carried by indirect methods. The direct observation of scattering by dust-free air followed considerably later. It was made independently by Cabannes,[1] by Schmoluchowski,[2] and by the present writer.[3] Previous workers (*e.g.* Tyndall) who had attempted the problem were discouraged by observing that the track of a powerful beam through air became apparently dark when means were taken to remove the dust by filtration. It must be remembered, however, that the effect to be expected is very faint. The thickness of air illuminated in any laboratory experiment can only be a small fraction of the height of the homogeneous atmosphere. Thus if a beam of sunlight 5 in. square were used, the intensity seen laterally should be only about $\frac{1}{50000}$th of the brightness of the sky.

Success depends not so much on increasing the intensity of illumination, or using a beam of large diameter, for the limit of what is practicable in these directions is soon reached. The point to be attended to is rather to design the apparatus in such a way that no false light interferes with the observation. If this can be achieved, then the great sensitiveness of the eye or the cumulative action of the photographic plate can be brought to bear. The principle to be used is to examine the beam, whether visually or photographically, as it passes across the mouth of a black cave. If the cave is deep enough it gives an incomparably darker background than black velvet or any other blackened surface. The arrangement used by the writer in his earlier experiments is illustrated in *Fig.* 1. The beam enters by the window A. It is delimited by the diaphragm B. It passes out at C, and is viewed laterally by means of the window D. E is the black cave which forms the background. F (diagrammatic only) is a photographic lens.

When the apparatus is filled with the ordinary dusty air of the room, the track of the (horizontal) beam is strongly marked out by the illuminated dust. On drawing a current of filtered air through the apparatus the beam appears to casual observation to become invisible. But closer examination with a well-rested eye shows its track still marked out in a faint blue luminosity.[4] Moderate filtration with cotton-wool suffices to remove all dust from the air. The use of

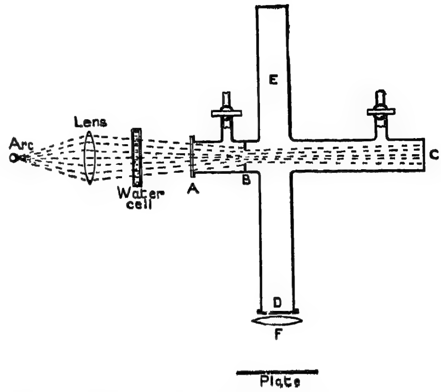


Fig. 1.—Diagram showing Arrangement of Apparatus.

a much longer and more tightly packed filter induces no further change.

§ (4) Polarisation and Intensity of the Scattered Light.—If a Nicol or double-image prism is held in front of the eye or the camera, it can readily be verified that the scattered light is almost completely polarised, the vibrations being vertical (beam horizontal) (see Plate, Nos. 5 and 6). A closer examination, however, shows that there is a slight residual defect of polarisation. The horizontal vibrations are too faint for visual detection, but can be brought out by prolonged photography. Experiments made in this way show that with air this faint component polarisation has about 4 per cent of the intensity of the strong one.

This result indicates much more complete polarisation than is found in the sky. It must be remembered, however, that that part of the sky which is at right angles to the sun is not illuminated by the sun only, but also by other parts of the sky. Moreover, the air usually contains dust particles which are not small compared with the wave-length of light. Both these causes tend to make the polarisation less complete.

As regards the slight defect of polarisation found in dust-free air, this is to be attributed to the non-spherical symmetry of the scattering molecules.

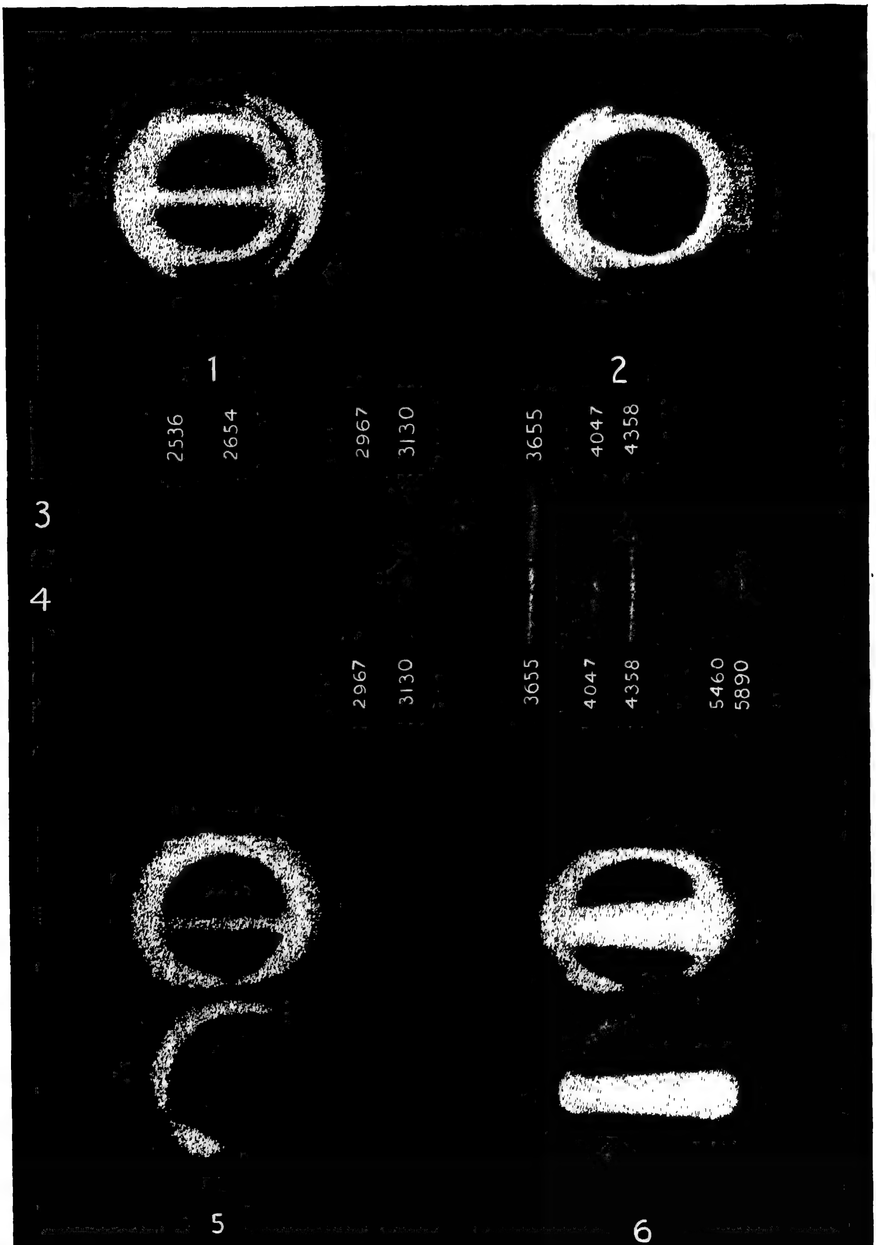
Similar incompleteness of polarisation is observed in other gases; the extreme cases

[1] *Comptes Rendus*, 1915, clx. 62; *Thèses présentées à la Faculté des Sciences de l'Université de Paris*, 1921.

[2] *Bull. Int. de l'Acad. de Sc. de Cracovie*, 1916, p. 218.

[3] R. J. Strutt (Rayleigh, 4th Baron), *Proc. Roy. Soc.*, 1918, xciv. 453; *ibid.* A, 1918, xcv. 155; *ibid.* A, 1920, xcvii. 436; *ibid.* xcviii. 57.

[4] Owing to a peculiarity of colour vision with faint lights, some persons are not able to recognise the colour as blue, though others feel no doubt about it. The use of colour-filters or a spectrographic examination, however, leaves no doubt that the scattered light is bluer than the incident (see description of Plate, Nos. 1, 2, 3, 4).



1. Beam in dust-free air viewed transversely. Ultra-violet filter. The oval outline is light diffused by walls of vessel. The beam is seen passing across this oval.
2. Similar conditions, except that a yellow filter is substituted. Beam very much fainter relative to light diffused by walls of vesse.
3. Spectrum of light from mercury lamp scattered by dust-free air and showing ultra-violet lines 2536 and 2654, but not the yellow and green lines.
4. Spectrum of mercury lamp direct, showing yellow line 5890 and green line 5460, but not the far ultra violet lines.
5. Beam in pure air through a double image prism. Vibrations in upper image vertical, in lower horizontal. The beam is invisible in the lower image, showing that polarisation of the scattered light is nearly complete.
6. Similar photograph with dusty air. The beams are very strong, and comparable in intensity

hitherto found are nitrous oxide, giving a weak component polarisation 15 per cent of the strong one, and argon, giving ·5 per cent for the same number. The nearly complete polarisation in argon will naturally be connected with the monatomic molecule. It is probable that data of this kind will afford a valuable clue in unlocking the secret of molecular and atomic structure. J. J. Thomson has discussed the subject from this point of view. The total intensity of scattering in a given gas is proportional to the density, and this relation has been tested in carbon dioxide at ordinary temperature up to the point of liquefaction. There is no evidence of molecular aggregation, which would result in increased scattering, up to this point. The proportionality to density affords an experimental proof that the intensities due to the separate molecules are additive, and consequently have their phases distributed at random.

The total intensity of scattering in different gases has been compared, with the result that it is nearly proportional to the square of the refractivity. This is in accordance with theoretical anticipation for a spherical molecule, an assumption which for this particular purpose seems to represent the main features of the case.

§ (5) Resonance Effects.—All that has been said so far applies to cases where the period of the incident light is very different from the free period of the gaseous molecules or atoms. With ordinary gases this condition is fulfilled, since the free periods are in the remote infra-red or ultra-violet. But in certain cases of metallic vapours it is possible to examine phenomena of resonance. R. W. Wood [1] has shown that when mercury vapour at the ordinary temperature is illuminated by light of the ultra-violet line in the emission spectrum of mercury at $\lambda 2536$, a copious scattering occurs. The scattered light is but slightly polarised, and the transition between this case and the ordinary scattering, in the absence of resonance, has not been fully traced on account of experimental difficulties.

Wood has also observed resonance in sodium vapour illuminated by D light.[2] The present writer extended the observation to the light of the ultra-violet line $\lambda 3303$, which is the member of the principal series next to the D line.[3] He also showed that illumination by $\lambda 3303$ gave rise to a secondary emission of the D line. This observation, however, belongs more properly to the subject of fluorescence.

R.

[1] *Phil. Mag.*, 1912, xxiii. 689.
[2] *Ibid.* 1905, x. 513; also *Physical Optics*, 1912.
[3] R. J. Strutt (Rayleigh, 4th Baron), *Proc. Roy Soc.* A, 1919, xcvi. 272.

Scattering of Light by Small Particles, Rayleigh's law of. See "Ultramicroscope and its Applications," § (2); also "Scattering of Light by Gases."

Schmidt and Haensch's Polarimeter. See "Polarimetry," § (13) (iii.).

Schumann Waves: a name given to waves of extremely short length measured by the use of a vacuum spectrograph, fluorite prism, and special photographic plates. See "Wave-lengths, The Measurement of," § (6).

Screen, Kinematograph, surfaces suitable for. See "Kinematograph," § (11).

Screen, Photometer: the surface which receives the light from the sources being compared. See "Photometry and Illumination," § (53).

Screws, manufacture and testing of precision, as used in optical measurements. See "Diffraction Gratings, Manufacture and Testing of."

Search-light Mirrors, The Silvering of, by Electrical Deposition. See "Silvered Mirrors and Silvering," § (3) (i.).

Search-light Projectors, Photometry of. See "Photometry and Illumination," § (115).

Seasonal Variation of Daylight. See "Photometry and Illumination," § (74).

Secondary Spectrum: the term given to the residual colour effects in a simple achromatic optical system. See "Microscope, Optics of the," § 17; also "Telescopes," § (9).

Sector Disc: a device for reducing in a known ratio the illumination on one side of a photometer head. See "Photometry and Illumination," § (21).

"Seed," a Defect in Glass: fine bubbles which should be removed during the process of fining, but which are sometimes found in the finished glass. See "Glass," § (16) (iii.).

Seidel, von, The Five Aberrations of. See "Telescope," § (3); also "Lens Systems, Aberrations of."

Selenium Cell as a Physical Photometer. See "Photometry and Illumination," § (33).

Semi-indirect Lighting. See "Photometry and Illumination," § (71).

Sextant. See "Navigation and Navigational Instruments," §§ (19), (20), (21).

Sextants using bubbles and pendulums. See "Navigation and Navigational Instruments," § (21); also "Aircraft Instruments," § (9), Vol. V.

Sharp Foot-candle Meter: a portable illumination gauge. See "Photometry and Illumination," § (63).

Sharp-Millar Illumination Photometer. See "Photometry and Illumination," § (59).

SHELL-BURSTS, LOCATION OF. See "Sound Ranging," § (5).

SHUTTER, EXPOSING: a mechanical contrivance for exposing the sensitive plate in the camera. See "Photographic Apparatus," § (8).

SHUTTER, SPEED OF, measurement of, by various methods. See "Shutters, Testing of Photographic," § (2).

SHUTTERS, TESTING OF PHOTOGRAPHIC

§ (1) INTRODUCTION.—The problem of accurately measuring the speeds of camera shutters has engaged the attention of a great number of investigators, but the methods evolved, though differing greatly as regards the nature and complexity of apparatus employed, are all based on the measurement of small intervals of time.

The term "speed" really denotes the total duration of exposure at any point of the plate. In the case of between-lens sector shutters this is equivalent to the total time during which the shutter leaves are open. In the case of focal plane shutters the speed is usually specified as the time taken for the opening of the blind to pass a point in its own plane opposite the centre of the plate. The duration of exposure at any other point usually has a somewhat different value owing to the variable velocity with which the blind moves. These durations in the plane of the blind differ from the intervals for which the corresponding points of the plate itself are exposed.

The speeds of ordinary shutters vary from about 1 to 1/1000 second, though some types of shutter are designed to give speeds of the order of 1 to 5 seconds. In order, therefore, that a method should give accurate results over the ordinary range of speeds it should be capable of measuring intervals of time to an accuracy of from 1/1000 to 1/10000 second. It should be noted in this connection that for most purposes it is not necessary to obtain an accuracy of, say, one per cent, since in most shutters the speed corresponding to any given setting varies considerably in successive exposures.

§ (2) MEASUREMENT OF SPEED. (i.) *Simple Methods.*—Before describing the more accurate methods of measuring shutter speeds it may be of interest to refer to one or two simple methods which do not involve very complicated apparatus. One method consists of forming an image of an illuminated pinhole by means of a lens and placing the shutter to be tested in the path of the beam of light. If, now, a photographic plate or a piece of sensitive paper is made to move with a constant definite velocity in the plane which is normal to the optical axis of the lens system and passes through the point image, a line is formed on the plate or paper, when the shutter is released. The length of the line gives a measure of the speed, provided that the velocity of the plate or paper is known. Instead of causing the plate to move in a plane it is perhaps more convenient to use a piece of sensitive paper wound round a drum, which is made to rotate at a constant rate, the axis of the drum being normal to the optical axis of the illuminating system. The accuracy of this method depends, of course, on the constancy with which the sensitive surface is moved and on the accuracy with which the motion can be determined.

A method employed some years ago consisted in photographing a bright ball forming the bob of a pendulum,[1] the ball being released from a known height and photographed in its first oscillation. A black background, with white lines drawn at distances representing equal intervals of time, was used for determining the speed of the shutter, this being given by the length of the trace of the bob image.

A simple instrument for testing shutters has recently been put on the market by a German firm under the name "Columbus."[2] It consists of a tuning-fork, giving 50 complete vibrations per second, between the prongs of which there is a small electric bulb. A sharp image of the lamp filament is formed by a microscope objective attached to one of the prongs of the tuning-fork. The instrument is set up in front of the camera whose shutter is to be tested, so that, when the fork is at rest, the image of the filament is focussed on the ground glass. A horizontal slit is inserted in front of the camera objective; when the fork is caused to vibrate, the illuminated point image is drawn out into a horizontal line. If, now, an exposure is made while the dark slide is moved downwards in the camera, a wavy line is obtained. The speed of the shutter is then determined from the number of waves formed. The instrument enables one to measure the speed for the central rays and also the difference in speed for the central and marginal rays.

Useful information on simple methods of shutter testing is to be found in recent volumes of the *British Journal of Photography*, in the German work L. David's *Praktikum*, 3rd edition, and in a paper by P. Schrott in the *Photographische Korrespondenz*, Oct. 1919.

(ii.) *More accurate Methods.*—

For more accurate determinations of shutter speeds a number of devices, such as rotating discs provided with spokes, vibrating singing flames, and tuning-forks, have been employed for giving the necessary short intervals of time. The first-named arrange-

[1] For discussions on pendulum methods of measuring shutter speeds see *Brit. Journ. of Photography*, 1906, liii.

[2] *Zeits. f. Feinmechanik*, 1920, xxviii. 58.

ment was used by Abney,[1] the spokes which project from a rotating disc being made to interrupt the beam of light which passes through the shutter. The speed of rotation of the disc is determined by means of the note given out by a siren arrangement.

Methods of measuring shutter speeds by utilising the motion of a revolving disc, containing a slit or slits, across a stationary illuminated slit, have been described by A. Kershaw[2] and J. de G. Hunter.[3] The former varies the speed of rotation of the disc until, on looking through the aperture of a shutter, the stationary slit appears to have no dark portion, indicating that the speed of the shutter is the same as the period of the rotating slit. In the case of focal plane shutters the stationary slit is arranged perpendicular to the direction of motion of the blind. Hunter's method is somewhat similar, except that he measured the length of the trace transmitted by the stationary and rotating slits. As the positions of the first and last points of the trace have to be determined, a number of exposures are necessary at each speed, and in order to keep the trace at the same position an automatic releasing device is used so that the exposure is made when one of the slits in the rotating disc comes to a certain position. The method has the advantage that photographic records can be made at the different exposures. The shutter speed is determined from the length of the trace and the speed of rotation of the disc.

Benoist[4] has utilised the vibrations of an acetylene gas flame to give equally spaced short intervals of time. An image of the flame is formed on the plate of a camera. When this is swung round during the exposures, a series of images is formed and the speed of the shutter can be determined if the frequency of the vibrations of the flame is known. One form of apparatus employed in this method is as follows. A box is provided on one face with a small burner, at the extremity of which the acetylene led into the box is lighted. The opposite face, which is formed by a rubber membrane, is fixed against an opening arranged at a convenient height in the wall of an organ pipe of known frequency. When the organ pipe vibrates under the action of suitable bellows, the flame vibrates with the same frequency. A somewhat simpler arrangement is to use a flame of acetylene emitted from the end of a pointed tube which is surrounded by a glass chimney. Such a flame "sings" and the note can be varied by altering the length of the chimney. The frequency of vibration of the flame can be determined by means of the note. One disadvantage of the singing flame method is that the trace obtained is a discontinuous one, so that the shutter speed can only be determined with an accuracy equal to the time interval of one vibration of the flame. Benoist[5] also suggests the employment of the vibrations of an electric arc run on alternating current of known frequency.

A method employed by the Eastman Kodak Company[6] consists in taking a number of photographs of the shutter during its period of operation, the time of exposure of each photograph being small in comparison with the rate of movement of the shutter. A beam of light from an arc lamp in a small projection lantern is reflected from a lightly mounted aluminium crown on which are placed 20 small plane mirrors held vertically on the face of the crown; the lantern condenser focusses an image of the arc crater at the mirror surface. The crown of mirrors is rotated about a vertical axis at a speed of 50 revolutions per second by means of a motor controlled by a centrifugal governor. The beam of light is thus interrupted 1000 times per second, a frequency which has been found most suitable for general testing. The beam is limited at the condenser by a vertical slit 2 mm. wide, the width of the beam as it flashes by reflection past the shutter opening being about 1/30 of the distance between the flashes; thus each exposure is about 1/30000 second. The reflected beam falls on a simple lens behind which the shutter to be tested is held, and a small camera lens of 90 mm. focal length forms an image of the shutter on the rim of an aluminium wheel 12 inches in diameter, around which a band of negative cinematograph film is fastened. This wheel is then rotated about a vertical axis by means of a crank and gearing at such a speed as to separate the different images of the shutter. Both the speeds and the efficiencies of a shutter may be determined by this method. For recording shutter speeds of 1/10, 1/5, and 1/2 second, the image of the shutter opening is restricted to a narrow band by inserting a 1 mm. slit in front of the box in which the moving film is enclosed. In order to save counting the hundreds of images obtained at these speeds one of the twenty mirrors may be painted black.

A very accurate method of testing shutter speeds, depending on the use of tuning-forks and vibration galvanometers, has been described by Campbell and Smith.[7] A plan of the general arrangement of apparatus employed in a modified form of the experiment is shown diagrammatically in *Fig.* 1. The light from a "Pointolite" tungsten arc

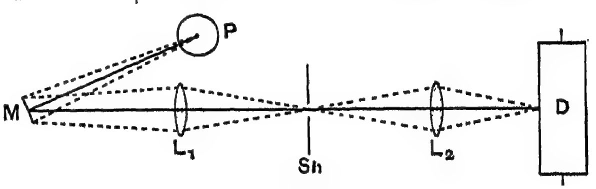


FIG. 1.

P is reflected from the mirror M of a vibration galvanometer and then passes through two lenses L_1 and L_2. A diminished point image of the incandescent tungsten ball is thus formed on the surface of a drum D which

[1] Sir W Abney, *Treatise on Photography*, 1905, 290.

[2] A. Kershaw, Patent No. 16,053, 1904.

[3] J. de G. Hunter, *Opt. Soc. Trans.*, 1906, viii. 1.

[4] L. Benoist, *Brit. Journ. of Phot.*, 1911, lviii. 949.

[5] L. Benoist, *Bull. Soc. Fr. Phot.*, 1910, 390.

[6] P. G. Nutting, *Brit. Journ. of Phot.*, 1916, lxiii. 394.

[7] A. Campbell and T. Smith, *Phys. Soc. Proc.*, 1907-9, xxi. 788; *N.P.L. Collected Researches*, 1910, vi. 93; see also T. Smith, *Opt. Soc. Trans.*, 1910-11, xii. 128.

can be rotated about a horizontal axis at any convenient speed by means of a small-power motor with variable gearing. The shutter *Sh*, whose speed is to be measured, is placed in the position where an image of the tungsten ball is formed by the first lens L_1. A piece of sensitive bromide paper is wound round the drum. When the vibration galvanometer, which is in tune with an electrically driven

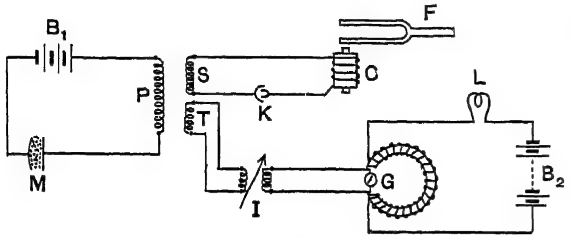


FIG. 2.

tuning-fork of known frequency, is set in action, the mirror M vibrates about a vertical axis, so that the point image on the surface of the drum is caused to vibrate in a horizontal plane. A wavy image is thus imprinted on the bromide paper when the drum is rotated. The number of waves formed depends on the total duration of exposure of the shutter and the frequency of the tuning-fork employed. When measuring the time trace which is obtained, it is most convenient to count the number of complete half-waves and estimate the end portions in terms of their projections on a line parallel to the axis of rotation of the drum. In order to get the greatest possible accuracy in determining the time intervals corresponding to these portions, it is necessary to remember that the motion of the point image is simple harmonic. By means of an easily calculated conversion table the lengths of the above-mentioned projections, expressed as fractions of the amplitude of the wave, may be obtained in terms of time intervals. In measuring the speed of a focal plane shutter a fine horizontal slit is mounted at the centre of the shutter and as near to the blind as possible; it is placed at the position where an image of the tungsten ball is formed by the lens L_1 (*Fig.* 1). In the case of speeds where the width of the slit is not small compared with the width of the blind opening, a small correction has to be applied to the result obtained, if great accuracy is required.

For testing speeds over the whole range commonly met with, it is convenient to use a number of tuning-forks giving, say, 50, 250, and 1000 vibrations per second respectively.[1] Each of these is electrically maintained in vibration and is in tune with a vibration galvanometer of corresponding frequency.

The electrical circuits employed are indicated diagrammatically in *Fig.* 2.

The primary coil P of a transformer is connected through a microphone hummer [2] M to a battery B_1 of from 6 to 8 volts. The microphone is mounted on or near the tuning-fork (or bar) so as to be set in vibration with it. The current, which is induced in the secondary coil S of the transformer by the intermittent current in the primary circuit, passes through a condenser K and a coil C which attracts one of the prongs of the tuning-fork F. A tertiary coil T of the transformer is connected through a variable mutual inductance I to the moving coil of the vibration galvanometer G; [3] the amplitude of vibration of the galvanometer mirror can be varied by altering the mutual inductance. The field of the galvanometer is maintained by the current from a battery B_2 of 100 volts, a lamp resistance L being included in the circuit. In order to get the best results the following condition should be satisfied as nearly as possible:

$$LK\omega^2=1,$$

where $\omega=2\pi n$,
L=coefficient of self induction of the secondary circuit,
K=capacity of condenser,
and n=frequency.

The method of maintaining tuning-forks in vibration by means of microphone hummers gives a good deal of trouble in practice. It is being superseded

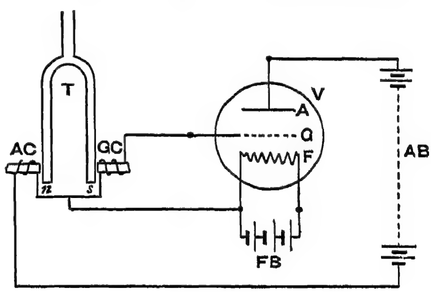


FIG. 3.

by a method which has been made possible by the recent development of triode valves in connection with wireless telegraphy.[4] The method is shown in *Fig.* 3.

[1] A vibrating cylindrical bar having a frequency of 1000 per second may be used.

[2] For information *re* microphone hummers see A. Campbell, *Roy. Soc. Proc.* A, 1906, lxxviii. 208; *N.P.L. Collected Researches*, ii. 268.

[3] A. Campbell, *Phys. Soc. Proc.*, 1905 7, xx. 626; *Phil. Mag.*, 1907, xiv. 494; *N.P.L. Collected Researches*, 1908, iv. 201.

[4] W. H. Eccles, *Phys. Soc. Proc.*, 1919, xxxi. 269; W. H. Eccles and F. W. Jordan, *Electrician*, 1919, lxxxii. 704; for theory, see S. Butterworth, *Phys. Soc. Proc.*, 1920, xxxii. 345.

The filament F of the triode valve V is heated by the battery FB. The anode A is connected through the anode battery AB and the coil AC to one terminal of the filament, and the grid G is connected to the filament through the coil GC. The tuning-fork T is permanently magnetised by an auxiliary magnet, on whose pole-pieces the coils AC and GC are wound, the poles of the fork being indicated by *n* and *s*. Suppose the fork to have been set in vibration, and consider the moment when the two prongs are moving away from their respective coils. The motion of the pole *s* will induce an E.M.F. in the coil GC which will raise the potential of the grid and so increase the current flowing into the valve at the anode. This current leaving by the filament completes its circuit through the coil AC, round which it passes in such a direction as to repel the pole *n*, and thus assist its motion. The energy necessary for keeping the vibration galvanometer in action may be tapped off from a coil placed near AC.

§ (3) Measurement of Efficiency.—In addition to the determination of speed it is sometimes necessary to measure the efficiency of a shutter at its different speeds. The method employed in the case of sector shutters is shown in *Fig.* 4. A time trace is formed on sensitive bromide paper, wound round a rotating drum D, by means of a system of lenses L_1 and L_2, as in the case of the speed test. In addition a horizontal slit S is placed in a diametral plane of the shutter *Sh* and as close to the shutter leaves as possible. This slit is illuminated by light from the same "Pointolite" lamp P as is used for the time trace, a cylindrical lens L_3 being employed to give a drawn-out image of the tungsten ball. An image of the slit is formed on the surface of the drum by means of a lens L_4; it is advisable to arrange that the images of the slit and the tungsten ball should lie on the same horizontal line. If, now, the shutter is released while the drum is rotating, a record is obtained, giving at each instant a measure of the length of slit uncovered. Such a record for a sector shutter is shown diagrammatically in *Fig.* 5.[1] The time trace may be limited to one complete revolution of the drum by inserting an auxiliary shutter, set at the necessary speed, between the lenses L_1 and L_2; the two shutters can then be released at practically the same instant. A series of measurements of the area of the shutter aperture, corresponding to different lengths of the slit opening, is then made by projecting an image of the former on to a piece of sensitive bromide paper, a fine thread being stretched across the shutter at the position occupied by the slit. Such a series of images is illustrated in *Fig.* 6, the white line in each case representing the position of the thread. From records like those shown in *Figs.* 5 and 6 a curve is constructed, giving at each instant of the exposures the area of the shutter aperture through which light passes.

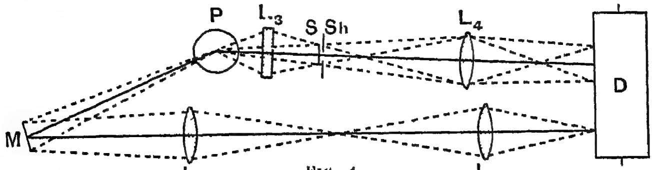


Fig. 4.

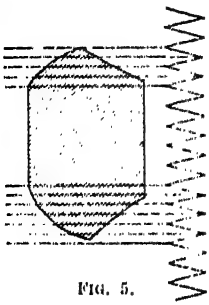


Fig. 5.

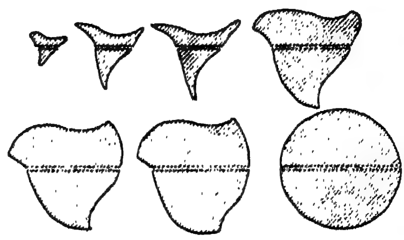


Fig. 6.

Now, if τ = total duration of exposure,
T = equivalent exposure,
a = area of shutter aperture at time t,
and A = maximum area of shutter opening,
it follows that

$$AT = \int_0^\tau a\,dt.$$

Thus the efficiency $= \dfrac{T}{\tau} = \dfrac{\int_0^\tau a\,dt}{A\tau}$.

[1] It is sometimes preferable to form the time trace across the centre of the efficiency record.

The value of the numerator is obtained by measuring the area bounded by the above-mentioned curve and the time axis. Hence the efficiency may be determined.

The efficiency of a focal plane shutter may be obtained theoretically from a consideration

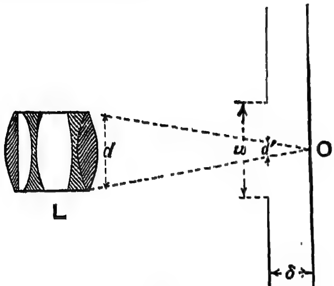


FIG. 7.

of *Fig.* 7. Let L be the lens and P the plate of a camera, B representing the roller blind. Let

w=width of opening in blind,
v=velocity of blind,
δ=distance between blind and plate,
f=focal length of lens,
d=diameter of lens aperture.

Then, using the same notation as before, the total duration of exposure τ at a point O is given by

$$\tau=\frac{w+d'}{v},$$

where d' is the diameter of the beam of light which the lens focusses at O, measured in the plane of the blind.

But $$d'=\frac{\delta}{f}\cdot d,$$

therefore $$\tau=\frac{w+\frac{d}{f}\delta}{v}.$$

On the other hand, the equivalent exposure T at O is given by

$$T=\frac{w}{v}.$$

Thus the efficiency $=\frac{T}{\tau}=\frac{w}{w+\frac{d}{f}\delta}$.

In focal plane work with moving objects the user is interested in τ, while the maker reckons to give T when he does not rely on his imagination.

By way of example we may take $v=100$ inches per sec. (a higher value than usual), for $T=1/1000$ sec., in which case $w=\cdot 1$ inch. Then, if $d=f/4$ and $\delta=\cdot 5$ inch, $\tau=9/4000$ sec.

With a more usual value of v, say 50 inches per second, $T=1/1000$ sec. corresponds to $\tau=7/2000$ sec. Thus the nominal superiority of the focal plane shutter for high speed work is largely one of nomenclature, and the difference between such shutters and good between lens shutters is less than the figures would indicate.

The method of measuring the efficiencies of sector shutters has been employed by the writer for obtaining information with regard to the motion of the blind in focal plane shutters. The shutter to be tested is mounted at S*h* (*Fig.* 4) in such a way that the length of the blind opening is at right angles to the length of the slit S. When the shutter is released, two traces are obtained on a piece of sensitive bromide paper wound round the drum D. *Fig.* 8 represents diagrammatically

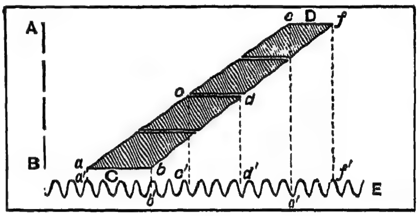


FIG. 8.

the type of record that would be obtained, on unwinding and developing the paper, in the case of a blind moving with uniform acceleration. AB is the image of the slit when the drum is at rest and the shutter is fully open; the clear spaces represent the images of fine wires stretched across the slit at certain points for the purpose of reference. A number of interesting data can be obtained from a study of the trace (CD). Let lines be drawn through the points a, b, c, d, e, and f parallel to AB, so as to cut the time trace E in the points a', b', c', d', e', and f'. Then, if we neglect the finite width of the slit image and also the distance of the slit from the shutter blind, we obtain the following results:

(i.) a' b', c' d', e' f' represent the times taken by the blind opening to pass certain points such as the top, middle, and bottom of the plate.

(ii.) a' e' and b' f' give the total times taken by the edges of the blind opening to traverse the plate.

(iii.) The slopes of the edges of the trace (CD) at any point give measures of the velocities with which the edges of the blind opening pass that point.

J. S. A.

SILICA, as used in the manufacture of glass. See "Glass," § (4).

SILVERED MIRRORS AND SILVERING

§ (1) Historical.—Polished silver surfaces have long been used as mirrors on account of their great reflecting power. Looking-glasses made of polished silver were used by the ancients, and in more recent times similar surfaces were given parabolic forms for use as reflectors in lighthouses. Such polished surfaces were capable of reflecting about 60 per cent of the incident light, but on account of the rapid loss in reflective power, due to tarnishing, it was more common to use polished surfaces of speculum metal, which does not tarnish readily and reflects about 63 per cent of light.

The advantages of a combination of transparent glass backed by a highly reflective medium are so great that mirrors were made by pressing mercury between glass and tin foil, the two metals forming an amalgam which adhered to the glass and in time hardened. This process is in use to-day and is fairly satisfactory for flat surfaces, but it requires considerable skill to secure a surface free from blemishes; and, moreover, it is not suitable for surfaces with curvature. The process is often called "soft silvering."

The chemical deposition of metallic silver on glass as now practised probably owes its origin to Liebig,[1] who found in 1835 that a brilliant deposit of metallic silver resulted from heating an aldehyde with an ammoniacal solution of nitrate of silver in a glass vessel.

It seems, however, that a process of silvering glass without quicksilver was practised by Thomas Rogers of London as early as 1750, but no details of the method are available.[2]

Silvered glass spherical mirrors, for lighthouses, were made by Letourneaux before 1840, and the idea of silvering glass mirrors on the *face* for use in telescopes appears to have occurred to both Steinheil and Foucault independently.

Steinheil first silvered telescope mirrors in 1856, and refers to the "Liebig process," whilst Foucault's account, given in the *Comptes Rendus* of February 1857, refers to the "Drayton process." Drayton's are the first British patents on chemically deposited silver, being patents 9968/1843 and 12358/1848. Drayton used oil of cloves and grape sugar as reducing agents of an ammoniacal solution of nitrate of silver.

§ (2) Modern Methods.—From that time the records of the Patent Office show that there is scarcely an interval in which patents have not been taken out for new methods of chemically depositing silver on glass. Nevertheless, all the processes so far known are empirical ones, and there are so many conditions affecting the reaction that no hard-and-fast rules can be laid down. The most successful methods are given below.

The Rochelle salt process, introduced by Cimeg in 1861 (Patent No. 619), is almost universally used when the work is required to be silvered on the back; the "Brashear" or "Sugar" process is employed when the deposit is required on the face, as in telescope mirrors, and where the glass can be suspended face downwards in the silvering solution.

In silvering large areas of plate glass, the sheets are first washed with a solution of proto-chloride of tin and then levelled up on a warm table; the silvering solution is then poured over the prepared glass (being retained in position at the edges by capillary action) and allowed to stand until the deposit is sufficiently dense. Further applications of the solution can be applied to increase the density if necessary. The effect of the proto-chloride of tin solution is to produce a darker and more uniform deposit, and at the same time a deposit that is more adhesive and does not flake off. To procure a uniform and adherent deposit of silver it is necessary that the glass surface should be fresh. Surfaces of old standing, however chemically cleaned, do not silver well and have often to be reground and polished before a satisfactory deposit can be secured.

Plate glass known as "silvering quality," having received an extra finish, makes the best looking-glasses. *Sheet* glass usually makes mirrors exhibiting a yellow tinge, due to sulphiding from traces of sulphur acquired in the flattening *lier*; and, moreover, the deposit on a fine surface is rarely as adhesive as that on a worked one.

(i.) *Rochelle Salt Method.*—Two stock solutions are required—A and B.

Solution A: Silver nitrate, 10 grams; water, 80 c.c. When dissolved, add ammonia very slowly until precipitate is nearly redissolved. Dilute to 1000 c.c. with distilled water, and filter. Care must be taken to avoid an excess of ammonia, as on this the success of the silvering depends. Thus, before filtering, the solution should exhibit a faint brown coloration.

Solution B: Silver nitrate, 2 grams; dissolve in as little water as possible, and add 1000 c.c. of boiling water. To this, whilst boiling, add 1·66 grams of Rochelle salt (sodium potassium tartrate), and continue boiling for twenty minutes or more until the liquid is nearly clear, leaving a grey precipitate. Filter hot.

Equal parts of A and B are mixed, when silver immediately begins to deposit. Thus no time should be lost in pouring it over the surface to be silvered.

[1] *Annalen der Chemie und Pharmacie*, xcviii. 132.
[2] *Lighthouses*, Alan Stevenson, 1850, Part ii. 103.

In practice on a large scale the temperature is kept at 40° C., when a good film is deposited in about twenty minutes.

(ii.) *Brashear's or Sugar Method.*—In Brashear's process the most important thing is the sugar solution. This improves by keeping, a solution some months old being more effective than a new one. The composition of the reducing solution is as follows: 1000 c.c. distilled water, 100 grams loaf sugar; dissolve, add 100 c.c. alcohol and 5 c.c. nitric acid.

Solutions of 10 per cent of silver nitrate and of caustic potash are prepared separately, the latter one as wanted. A supply of ammonia and some very dilute ammonia are also required, the latter in order to obtain the pale brown colour of the ammoniated solution of silver nitrate that is necessary before adding the reducing agent.

Having selected a suitable dish to contain the liquid, in which the mirror can be placed face downwards with about ½ or ¾ in. of liquid underneath, find, on the basis of 1 of silver nitrate solution to 4 of the total required, the amount of silver solution needed. To this add ammonia till the first precipitate is dissolved; then add one-half of this quantity of the potash solution (this is a variation from Brashear's formula that works well); and again add ammonia till the mixed solution is quite clear, taking care to put only sufficient ammonia for that purpose; then add a weak solution of nitrate of silver till a clear brown colour is obtained. Should this become a dark brown, some of the weak solution of ammonia will bring it to a pale brown colour, which must persist.

The mirror, previously cleaned, is suspended in the dish in distilled water of sufficient amount to make up, on the addition of the solutions, the total liquid required; it is lifted out and the prepared solution mixed with the distilled water and an amount of the reducing solution equal to about one-half that of the nitrate of silver solution, more or less as the temperature is under or over 60° F.; the mirror is then immersed, beginning by dipping the edges first and lowering so as to prevent the formation of air-bubbles under the glass. The solution changes from pink to black, and in about thirty minutes sufficient silver is deposited.[1]

The thickness of the silver film suitable for telescopes has been found by Dr. Draper and Dr. Common to be about 0·0001 mm.

(iii.) *Alternative Methods.* — An alternative method to that of immersion for the application of the silvering solutions is one of spraying.[2] This is an economical process, as the two solutions are mixed almost at the moment of application. Usually the solutions are fed separately along the arms of a Y-tube, uniting in the third arm, which is under air pressure, the mixture issuing as a nebulous spray.

This method was tried, without much success, by Dr. Common in silvering large telescope mirrors, but in its modern application to searchlight mirrors silvered on the back, where the spraying is done much in the same way as paint is applied by the aerograph, and where the mirror can be rotated at about 100 revolutions per minute at the same time as it is sprayed, the process[3] has been quite successful in securing uniform films.

Mirrors silvered on the face become tarnished rather quickly unless protected from the action of traces of sulphides in the air; even then it is frequently necessary to re-polish or re-silver. With mirrors silvered on the back, protection may be given by coating the silver in various ways according to the subsequent treatment it is expected to receive. In ordinary looking-glasses the silver is first coated with shellac varnish or gold size, and subsequently given a backing of red oxide of iron paint. Other backings, containing bitumen or gutta-percha, are also used. With such protections the silver remains good almost indefinitely, defects only appearing through action of the glass on the silver, impurities left before silvering, or on exposure to the air by fracture of the backing.

§ (3) Electrical Deposition. (i.) *Searchlight Mirrors.*—Searchlight mirrors have metallic backings to protect the silver. First a coating of copper is deposited electrolytically on the silver, and frequently this receives a further backing of electro-deposited lead, which may be of sufficient thickness to form a support for the mirror in case of fracture. Such a series of layers, glass, silver, copper, and lead, is, on account of differences in coefficients of expansion, attended with disadvantages if suddenly heated, for the metallic backings expand beyond the elasticity limits of the glass and tear the silver from the glass surface, causing "blisters" to appear in the silver film. In some cases the silver adheres to the glass so intimately that a surface layer of the glass itself is torn away, producing a frosted appearance.

(ii.) *Metallic Mirrors.*—Advantage has been taken of the difference in the coefficients of expansion of metals and glass to produce metallic mirrors by electro-deposition.[4]

A glass surface accurately worked to the figure required receives a deposit of silver; this

[1] Further details of silvering glass are to be found in *The Observatory*, No. 193, 1892; *Ast. Phy. Jour.*, 1895, i. 252; *Nature*, Sept. 23, 1897, p. 505; *English Mechanic*, Feb. 3, 1911, Nov. 10 and 24, 1916; *Knowledge*, Nov. 1914, p. 402; *Conjoint Report on Silvering by the Physical and Optical Societies*, 1921.

[2] Patent 9977/1898, Barnes and another.
[3] Patent 127316/1917, Barnes.
[4] Patents 22730/1891, Lake, and 5600/1859, Cowper-Coles, 12005/1907, Cowper-Coles.

is then used as the cathode, and by electro-deposition of the required metal the mirror is built up to the necessary thickness. The whole is then subject to a sudden temperature change and the metal part sprung from the glass.

Silver cannot be electro-deposited on glass, except by first roughing the glass and preparing it with plumbago or making it conductive by other processes, *e.g.* burning gold or other metal into its surface and then using the prepared surface as the cathode.

(iii.) *Cathode Rays.*—A more recent application of the deposition of silver by electric means is by the use of cathode rays. In vacuum tubes provided with electrodes—*e.g.* X-ray tubes—prolonged discharge produces a blackening of the tube due to the deposit of a metallic film on the interior walls. In a similar way small glass surfaces have been coated with silver by this method of cathodic deposition or cathodic sputtering. The body to be coated is placed just outside the dark space surrounding the cathode and is connected to the anode. The cathode is disc-shaped or formed of a brush of fine wires, and the tube is hydrogen filled at a pressure less than 0·1 mm. mercury. Almost any metal is readily deposited on glass by this method of sputtering, and almost any degree of opaqueness can be secured.

(iv.) *Thermal Deposition.*—Thermal deposition of silver is practised by heating the metal in a small electric furnace in an exhausted vessel. The receiving surface is kept cool and brought close to the heated metal, when, as volatilisation takes place, it receives a deposit of the metal.

Mention should also be made of the method of spraying surfaces with silver. The surface to be coated is placed below an oxy-acetylene flame, into which is inserted silver wire. The glass travels to and fro under the wire or *vice versa*, and the force of the jet carries the volatilised metal on to the surface, where it condenses.[1]

§ (4) Dental Application.—One of the latest applications of silvering is in dentistry. Infected dental tissue is sterilised by treating with an ammoniacal solution of silver nitrate, which it readily absorbs. Healthy tissue is unaffected. The infected area is then isolated by a further application of formalin, which reduces the silver solution and deposits metallic silver. This fills up minute cavities, inaccessible by other means, and thus prevents further infection.

§ (5) Half-silvered Mirrors.—"Half-silvered" glass mirrors having a thin film of deposited silver have been applied to many kinds of optical instruments. Usually the film of silver is of such a density that half of the incident light is reflected and half transmitted. Such an arrangement is found in the Michelson interferometer[2] and also in a form of binocular microscope, where the light, after passing through the objective, falls on a half-silvered surface, the reflected and transmitted beams then being directed to the right and left eye respectively. Another application of the "half-silvered" surface is made in some forms of collimating gun-sights,[3] where the line of sight passes through the film to the target, whilst at the same time the eye sees, in the same optical direction, the reflection of a fiducial point contained in an attached collimator.

§ (6) Miscellaneous Applications.—To prevent loss of light in certain optical instruments it is often necessary to silver glass faces, as in prisms on which the light is incident at less than the critical angle; of these may be cited the end reflectors or pentagonal prisms, also the compound central prisms in use in many rangefinders.

A deposit of silver acting as an opaque screen has been made use of in the Abbe test plate, where a silver film on glass is ruled in such a manner as entirely to remove the metal along the rulings. The result, when seen under the microscope, is a coarse grating, with alternating opaque and transparent bars suitable for testing microscope objectives.

Next to the use of deposited silver for looking-glasses, probably the greatest commercial application is the silvering of the interior walls of vacuum flasks in order further to diminish the rate at which radiation occurs.

Glass and quartz fibres are silvered not for use as reflecting surfaces but to make them electrically conductive. In the Einthoven string galvanometer a silvered glass or quartz fibre is stretched in a strong magnetic field, the silvered fibre acting as a conductor and forming the moving system of the instrument.

§ (7) Amount of Reflection.—Steinheil, as long ago as 1857, found that freshly deposited silver reflected 91 per cent of the incident light at 45° incidence and reflection. More recently Hagen and Rubens[4] have determined the coefficient of reflection for perpendicular incidence and reflection all along the spectrum, from $\lambda 140000$ in the infra-red to $\lambda 2500$ in the ultra-violet, and find

[1] For fuller particulars see "Electrical and Thermal Disintegration of Metals," *The Chemical World*, ii. 140 *et seq.*

[2] *Smithsonian Contributions to Knowledge*, No. 842, 1892, p. 8 (footnote).

[3] "Collimating Telescope Gun-sights," Sir Howard Grubb, *Transactions Roy. Dublin Soc.*, March 20, 1901.

[4] *Ann. der Phys.*, 1900, i. 352, 1902, viii. 1, 1903, xi. 873.

that a face-silvered mirror reflects almost equally well all rays in the visible spectrum, the coefficients at λ7000 and λ4000 being 95 per cent and 85 per cent respectively. For values of reflectivity at varying angles of incidence see "Reflecting Power of Mirrors" by C. A. Chant.[1]

In the infra-red the coefficient of reflection rises to a steady 98 per cent, but there is a rapid falling off in the ultra-violet, where it diminishes to 4 per cent at λ3160, but rises to 34 per cent at λ2510. The trend of the curve, according to Hagen and Rubens, is upwards at this point, though Minor[2] finds a decreasing reflecting power beyond λ2500.

In consequence of this transmission of ultra-violet rays, silver reflectors are inadmissible in spectrographs and other instruments used for investigating the Schumann and extreme region of the ultra-violet.

Advantage has, however, been taken of this transmission band to use the silver deposit as a light filter for taking photographs in the almost monochromatic light around λ3000. (See "The Moon in Ultra-violet Light,"[3] and "Nickelled Glass Reflectors for Celestial Photography,"[4] by R. W. Wood.)

When we come to the reflectivity of glass mirrors silvered on the back—*e.g.* searchlight mirrors—there is a slight loss due to absorption and internal reflection in the glass. Measurements of the mean coefficient of reflection made at the N.P.L. in 1918 give a mean value of 88 per cent, whilst one of ten specimens attained the high figure of 94 per cent. The claim for the superiority of the gilded mirror over that of a silvered one has now been abandoned. The mean reflectivity of two such mirrors is about 70 per cent and 86 per cent respectively, but there are conditions of landscape which, under illumination from a gilded mirror, give greater contrast than when illuminated by a searchlight with a silvered mirror, owing to diminished reflection of the green rays by the gilded mirror. The same contrast effect, however, can be produced with the silvered mirror by making the observations through suitable light filters.

§ (8) Reflection of Polarised Light.—The phase change, when light is reflected in glass from a silver surface, depends upon the thickness of the silver, but for a thick film it appears to be three-quarters of a wave-length retardation.[5]

Circularly polarised light is changed to elliptically polarised by reflection from silver surfaces. Details of the change in position of the axis of the ellipse due to changes of incidence are given in a "Note on the Polarising Effect of Coelostat Mirrors."[6]

W. S.

[1] *Ast. Phy. Jour.*, 1905, xxi. 211.
[2] *Ann. der Phys.*, 1903, x. 621.
[3] *Mont. Not. R.A.S.*, 1910, lxx. 226.
[4] *Ast. Phy. Jour.*, 1911, xxxiv. 404.
[5] *Nature* (Edser and Stansfield), 1897, lvi. 505.

Simmance-Abady Flicker Photometer. See "Photometry and Illumination," § (98).

Sine Condition: a relationship which must be fulfilled for rays passing through different zones of an optical system in order that the defect known as coma may be eliminated. See "Microscope, Optics of the," § (12); also "Telescope," § (3), and "Optical Calculations."

Single-ray Velocity, Axis of: a direction in a biaxial crystal in which a ray undergoes no double refraction; also known as a secondary optic axis. See "Polarised Light and its Applications," § (7) (ii.).

Siren: a device for producing a musical note of given frequency, in which a disc, with a circle of holes in it, spins so that the holes are alternately opened and closed by passing a similar set of holes in a fixed plate, an air-blast giving the sound; this provides a suitable method for evaluating the frequency of an organ-pipe. See "Sound," § (53) (ix.).

"Smashing" Point of a Lamp. See "Photometry and Illumination," § (78).

Sodium Oxide, Use of, in Glass Manufacture. See "Glass," § (5) (i.).

Solids, Specific Heats of, application of quantum theory to. See "Quantum Theory," § (5); also "Calorimetry, Quantum Theory," Vol. I.

Soprano, E♭: a brass wind-instrument of high compass. See "Sound," § (44).

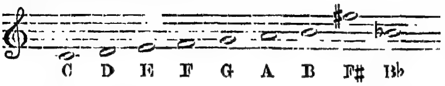
SOUND

1. Pitch, Scales, and Temperament

§ (1) Pitch and Frequency.—As the keys of a piano are struck in order from left to right, the various notes of the instrument are heard, and we pass, by a gradual transition, from the deep tones of the bass to the shriller tones of the treble. The sounds successively heard are said to be higher and *higher in pitch*, and pitch is that property of the note which determines its *position* in this musical range. It is found to depend solely on the number of vibrations per second occurring in that sound. This number of vibrations per second, or *frequency*, is accordingly adopted as a precise physical measure of the pitch. Musically, the pitch of a sound is denoted conventionally on the staff notation by the place of a certain symbol on a staff consisting of five horizontal lines and their associated spaces. The place of any staff among the pitches it is wished to

[6] *Ast. Phy. Jour.*, 1909, xxix 300.

denote is expressed by the clef at its commencement. Each such sound or note is called by one of the following letters, C, D, E, F, G, A, B. Further niceties are then expressed by the presence of a sharp (♯) or flat (♭), denoting a certain sharpening or flattening of that pitch indicated by the given position on the staff or its corresponding letter.[1]



§ (2) Interval.—When the above letters are used up the next musical sound to the right (higher in pitch) is called C again, sometimes distinguished by an accent (C′) to show that it occupies a higher place on the staff and in the sequence of possible notes. The frequency of this C′, when recognised by the musical ear as sustaining the true octave relation to C, may be shown experimentally to be just double that of C. Further, any two sounds, for which the ratio of their frequencies equals *two*, are found to have the same musical relation. They are accordingly said to be an octave apart. Or, in other words, the *interval* between them is said to be an octave. Similarly, with any other pairs of sounds which are recognised as sustaining some one definite musical relation, the *ratio* of their frequencies is a constant characteristic of and invariably associated with that relation and that only. Thus, the interval or range between any two musical sounds may be measured physically by the ratio of their frequencies, or by some other method closely linked with and dependent on this ratio. It thus follows that a large interval is measured by a ratio which is the product of those ratios which measure the small intervals of which the large one is composed. For, if three notes have frequencies in the relation 2, 3, and 4, the ratios of the small intervals are 3 : 2 and 4 : 3, while that of the whole or extreme interval is 4 : 2; and this is obviously the product of the other two. But this multiplication of ratios of component intervals to obtain that of the larger is sometimes inconvenient. For example, it does not lend itself to plotting intervals to scale, on a diagram, where the measure of the extreme interval must be the sum of those of the component intervals. To meet this need we take the logarithms of the ratios instead of the ratios themselves. Then, in the above numerical case, we have

$$\log 3/2 + \log 4/3 = (\log 3 - \log 2) + (\log 4 - \log 3) = \log 4/2.$$

That is, the logarithms *add* where the ratios multiply.

Musically, the intervals between pairs of notes are expressed by such terms as *second*, *third*, *fourth*, *fifth*, etc., corresponding to CD, CE, CF, CG, etc. To these terms various prefixes are used, such as *major*, *minor*, *diminished*, etc., to indicate the varieties possible in certain intervals though named by the same ordinal number. An alternative method for naming small intervals is the use of the terms tone (or whole tone) and semitone. Thus, the intervals CD, DE are tones, while EF is a semitone. It must not be taken to follow that CD is always exactly equal to DE, nor that EF is precisely the half of either of the above. The intervals EF, BC, AB♭ are called *diatonic* semitones; the intervals CC♯, FF♯, BB♭ being called *chromatic* semitones.

§ (3) Scales.—A scale is a series of notes proceeding up or down at certain specified intervals. These usually consist of tones and semitones, or of semitones only. Thus the following are examples of scales: (1) C, D, E, F, G, A, B, C; (2) C, D, E♭, F, G, A, B, C; (3) C, C♯, D, D♯, E, F, F♯, G, G♯, A, B♭, B, C. The first and second are diatonic scales, *i.e.* they include tones in their intervals. The third is a chromatic scale, *i.e.* it is coloured or affected by using notes not belonging to the diatonic scale. Further all the above are scales of C since they begin and finish on C, called the *keynote* of the scale, and have the notes appropriate to that scale. In the chromatic scale (3) all the intervals are called semitones (though not necessarily all equal). In the diatonic scales, (1) and (2), the intervals are either tones or semitones, but the order of their occurrence differs in the two scales. Thus in the first the order of intervals is: tone, tone, semitone, tone, tone, tone, semitone; while in the second we have as the order: tone, semitone, tone, tone, tone, tone, semitone. In the former case, the third from the keynote, CE, is called a major third and the scale a *major* scale. In the latter case, the third from the keynote, CE♭, is called a minor third and the scale a *minor* scale. These two forms of the diatonic scale are often referred to as the major and minor modes. The minor scale (2) is the ascending form of the melodic minor. The descending form would have B♭ and A♭ instead of B and A. There is also the harmonic minor which has the same notes in descending and ascending, viz.: C, D, E♭, F, G, A♭, B, C. Its intervals in ascending sequence are, therefore, tone, semitone, tone, tone, semitone, tone and half, semitone.

§ (4) Modulation.—Suppose that a piece of music began with the notes shown in scale (1) above and after a time the F was always replaced by F♯. It may then be found that the sequence of intervals is changed so that the scale in use is really that of G. For, on passing up from G, we have: tone, tone, semitone, tone, tone, tone, semitone; and this

[1] For production of sounds of various desired pitches see section "Sound Producers."

sequence is characteristic of the major diatonic scale. The piece of music is then said to have changed in key from C to G. If, on the other hand, when originally in the key of C major, after a time the E should be always replaced by E♭, the music will then have changed from the major to the minor mode, the key now being C minor. Such changes, either of key or mode, are called *modulations*.

§ (5) Intonation and Temperament.—By intonation is meant any precisely specified system of tuning the intervals in the various scales. A temperament is an intonation which, for the sake of simplicity, deliberately sacrifices some desirable exactness. The question naturally arises as to why this should become necessary. Let us examine this point carefully, having in view a keyboard instrument with notes of fixed pitches. An ideal intonation would secure three signal advantages, namely: (1) pure concords; (2) free modulation; (3) practical convenience. The first is needed to express truly the intentions of the composer and to satisfy critical ears, at any rate in slow music. The second is needed to enable the singers or players to take a piece in any desirable key and to allow of any unrestricted changes of key or mode which may occur within the piece itself. The third advantage is usually construed as being a limitation to twelve notes to the octave, though a few more may be allowable if arranged with studied simplicity. It must be remembered that the limitation of the number of notes in the octave is desirable for various reasons. Thus, if too great intricacy is entered upon, there is difficulty not only in tuning and retuning, but in playing and also in writing and reading the music.

Is there any inherent difficulty in simultaneously satisfying the above three requirements? To answer that we must find what they severally involve.

It is found by experiment that pure concords necessitate simple ratios of frequencies; that is, ratios expressible by small whole numbers. This has led to what is called *just* intonation, in which the frequencies of the notes are as follows. Any departure from these precise relations impairs the purity of the concords.

Relative frequencies	24	27	30	32	36	40	45	48
Notes	C	D	E	F	G	A	B	C′
Intervals as ratios	9/8	10/9	16/15	9/8	10/9	9/8	16/15	

In estimating the intervals by their ratios, that of the higher note to the lower is taken. The logarithms of these three sizes of intervals (which may be called the *large* tone, *small* tone, and semitone) are respectively 0·0511525, 0·0457575, and 0·0280287. Or, if we take numbers of four figures proportional to the logarithms, we have for the relative sizes of the three intervals of the just scale 5115, 4576, and 2803. Now, if we repeated these intervals throughout the scale as given above, we should have the large tone occurring three times, the small tone twice, and the semitone twice. This would give by addition a total number 30,103, to represent the interval of the whole octave from C to C′. Hence, if the notes in the scale which give concords with C are to occur at certain of the natural equal steps in a ladder throughout the octave, such a ladder would need over thirty thousand steps. And yet this, as we have seen, would be only an approximation to absolute truth.

But, for free modulation, we need to be able from any note to proceed by any interval of the scale. Further, all possible modulations must be linked up so that sharpened notes at some point or other coalesce with other flattened notes. Hence, for exactness, every step in the 30,103 would have to be provided, for no smaller numbers express the same ratios.

Thus, in the inevitable nature of things, the intonation which satisfies the first and second requirements fails utterly at the third requirement of practical convenience. Or we might equally say that what satisfies the first when limited to twelve notes in the octave, fails utterly to comply with the second.

Another and more practical way of regarding the matter is from the point of view of the violinist or violoncello player. Let two such players tune a certain string of each instrument to unison, and let one player sound that whole string while the other plays upon such a portion of that string as to give pure concords with the other. Thus, one might hold the keynote of a scale while the other in turn sounds the other notes of the scale. Let the exact *stops* (or positions of finger) on the string be marked. Then it would be found that they do not fall in with any simple subdivision of the octave into any small number of steps corresponding to equal intervals. Such equal intervals would be obtained on shortening the string by the same *fraction of its then sounding length* when proceeding to each new stop. Here lies the *root difficulty* of the whole problem. And it is *impossible* to remove it. We can but bow to it and do the best that is possible under the circumstances.

Yet another way of emphasising the difficulty and illustrating its inevitableness is as follows: Pass up from a given note, say 100 per second, in three ways, namely, by octaves, by fifths, and by major thirds. Then we obtain the following three series:

By octaves . .	100, 200, 400, 800, 1600, etc.
By fifths . .	100, 150, 225, 337½, etc.
By major thirds .	100, 125, 156¼, etc.

This at once shows that, in just intonation, no number of ascents by fifths or by major thirds will ever equal any other number of ascents by octaves, for the last immediately brings us to even hundreds, while the former two lead to odd numbers and fractions. In other words, the numbers concerned in these ratios (2, 3, and 5) are *prime* to one another.

It is therefore clear that every practical solution of the problem must be a compromise, some desirable exactness of concord being sacrificed for the sake of simplicity and convenience. And the way to find such practical solutions or temperaments is to seek some smaller number of steps in the octave such that certain exact numbers of them will give for each note of the scale a tolerable approximation to the concord sought. In all such cases the octave relation must be retained inviolate: for it is more sensitive to mistuning than any other interval (except the unison).

§ (6) THE CHIEF TEMPERAMENTS.—Such a system as described above is called a *temperament*, because in it certain intervals are purposely tempered or made to lose some accuracy of tuning in order to meet the requirements of practical convenience and allow of free modulation with fewer notes.

The only temperaments that have enjoyed any great vogue hitherto are the mean-tone and the equal temperament. But it seems desirable to notice briefly two others, viz. Bosanquet's Cycle of Fifty-three and Woolhouse's Cycle of Nineteen, as the comparison of the four places the whole subject in a clearer light.

(i.) *Bosanquet's Cycle of Fifty-three.*—This temperament is as near an approach to just intonation as could be demanded from any instrument with fixed keys. It retains a diatonic semitone exceeding the half of a tone, and even distinguishes between the large and small tones of the diatonic scale. As its name implies, it divides the octave into 53 equal parts or steps. It allots 9 of them to the large tone, 8 to the small tone, and 5 to the diatonic semitone. Its departures from the just intonation do not reach the hundredth of a tone, hence all the concords may be regarded as true. Its great fault lies here, *two* new notes are needed for each change of key to the next sharp or flat remove. This is due to the retention of the large and small tones (9 steps and 8 steps). Now from C to D the interval must be 9 steps for the key of C, but from G to A (for the key of C) the interval must be *8 steps*. Thus, when passing to the key of G, we require *9 steps* from G to A (in order to reproduce here the old interval C to D, between the keynote of a scale and its second note). Hence the first sharp remove from C to G involves *a slightly-sharpened A* as well as the usual F♯. Hence the first step towards simplicity is to drop the distinction between large and small tones as in the temperament to be noticed next.

(ii.) *Mean-tone.*—The *mean-tone*, as its name implies, uses one size of tone only, but makes it the *mean* of the large and small tones required by the just intonation. Thus, the interval C to E is made *quite true* on the mean-tone temperament, but is divided into two *equal* mean-tones. This puts the fifths (and fourths) slightly out of tune. For, as we have already seen, just major thirds never repeat so as to build up any number of just fifths. Also, since the fourth is the difference of an octave and a fifth, it is disturbed also. Still the harmony of the mean-tone was found to be wonderfully good and was preserved on organs for many years. It presented the signal advantage that only *one* new note was needed for each remove. The objection that drove it out of use was the fact that it still required so many notes to the octave for free modulation into any and every key. This was due to the retention of a diatonic semitone about three-fifths of the tone. Thus, a sharpened note really differed from that derived by flattening the note above. Hence to hear the mean-tone properly all these separate sharps and flats must be provided. But this was never done, its necessity being apparently only imperfectly understood. To realise the number of notes needed to render the mean-tone, we may note that Huygens's Cycle of Thirty-one is practically in agreement with it. Thus 31 notes to the octave are needed to give free modulation in the mean-tone, or in what is practically equivalent to it. This cycle allots 5 steps to each tone and 3 steps to the diatonic semitone, leaving 2 steps for a chromatic semitone. When organs with only five black digitals to the octave were tuned to mean-tone, only five keys could be played in beside the natural key of C. When the attempt to play in other keys was insisted upon, then notes tuned as sharps were used when flats were really needed, or *vice versa*, and disaster followed. But thirty-one notes to the octave are too many to expect on an organ and were probably never provided. Hence the mean-tone has not been really offered along with free modulation.

(iii.) *Woolhouse's Cycle of Nineteen.*—The next thing is obviously to reduce the number of steps in the octave while retaining, if possible, the diatonic semitone greater than half a tone as it should be. This was done in the proposal of Woolhouse to divide the octave into 19 steps, allotting 3 steps to each tone and 2 steps to each diatonic semitone. The first question to be asked is as to the approximation to just intonation which this cycle offers as regards the thirds and fifths. It is found to be nearly as good as the mean-tone, having true minor thirds and major

sixths, the fifths and fourths being a little out of tune. It is to be specially noticed that the sharps and flats of this temperament differ not in name only but in tuning. Further, that with nineteen notes in the octave modulation is quite unrestricted. The notes are

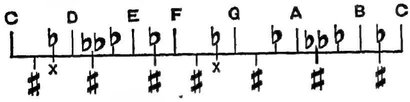


FIG. 1.—Woolhouse's Cycle of Nineteen.

shown to scale in *Fig.* 1. With this cycle it would be possible to play in the extreme keys B♭♭, F♭, G♯, and D♯ beside the usual 15 keys. For slow sustained music it should be easy enough to play on a keyboard arranged to sound this cycle. The present black digitals could be made with a front half (dull red, say) for the sharps and a back half (dull green, say) for the flats, then two shorter black digitals could be placed between E and F and between B and C. They would be the sharp of the lower note and the flat of the upper.

(iv.) *Equal Temperament.*—This familiar temperament makes the maximum sacrifice of harmony for the sake of free modulation with the minimum number of notes to the octave, viz. 12. The semitones, whether diatonic or chromatic, are just half the tones, hence only 12 steps are needed. This temperament makes the intervals of the major third and major sixth much too large, those of the minor third and minor sixth much too small. Hence the common chords in either mode are far from ideal.

(v.) *Comparison of Various Temperaments with Just Intonation.*—Much of what has preceded on the different temperaments is compactly exhibited in the accompanying Table I.

TABLE I

CHIEF MUSICAL TEMPERAMENTS AND JUST INTONATION

Names and Steps.	Scales, Intervals, Steps, and Errors.								Concords.	Keys.	Notes.
	C.	D.	E.	F.	G.	A	B.	C.			
Just—										1	7
Vibrations	24	27	30	32	36	40	45	48	All true	3	11
Intervals	0	1·02	1·93	2·40	3·51	4·42	5·44	6·00		12	20
										Free modulation	Unlimited number
Bosanquet's	0	1·02	1·925	2·49	3·51	4·415	5·434	6·00	All practically true	3	11
9, 8, 5, 9, 8, 9, 5 =53 steps	0	9	17	22	31	39	48	53		6	17
Errors	..	..	0·5♭	..	..	0·5♭	0·6♭	..		12	25*
										23	31* (51 in name)
										53*	53
Huygens's (mean-tone nearly)	0	0·965	1·93	2·52	3·48	4·45	5·41	6·00	True major third	6	12
5, 5, 3, 5, 5, 5, 3 =31 steps	0	5	10	13	18	23	28	31	Very good minor third	11	17
Errors	..	5♭	..	3♯	3♭	3♯	2♭	..		24	30
										31*	31*
	..	..	..	..	..	..	..	..	Rather flat major third and fifth	6	12
										11	17
Woolhouse's	0	0·95	1·895	2·526	3·474	4·42	5·37	6	True minor third	12	18
3, 3, 2, 3, 3, 3, 2 =19 steps	0	3	6	8	11	14	17	19	Second between D and 'D	19*	19
Errors	..	7	3·5♭	3·6♯	3·6♭	..	7♭	..			
Equal	0	1	2	2·5	3·5	4·5	5·5	6	Very sharp major third sixth, and seventh	12	12
										or 15*	12
2, 2, 1, 2, 2, 2, 1 =12 steps	0	2	4	5	7	9	11	12		or 20*	12
Errors	..	2♭	7♯	1♯	1♭	8♯	6♯	..			

* Fewer notes required for given number of keys than might have been expected. This is because in the full number of notes by *name* some are identical in pitch.

For the just intonation the numbers of vibrations in the same time are given for the various notes of the scale. Then below follow the intervals up from C expressed to the nearest hundredths of the equal temperament tones. For the four temperaments which follow, the vibrations are not given, but the intervals up from C are expressed both in equal tones and in the steps characteristic of the temperament in question. In a third line occur the errors of the temperament from just intonation expressed in hundredths of an equal tone. To give an idea what sort of interval this hundredth of a tone is, it may be noted it is that between lengths 870 and 869 of a given stretched string. This may be realised in millimetres on a monochord. Again, to make clear what the errors of a temperament mean when expressed as a change of stop on an instrument of the violin family, take Woolhouse's cycle and consider a violin string of vibrating length 13 inches (from bridge to nut). Then, if the open G string be taken as the keynote of a scale which is to be played on that string alone, the lengths of vibrating parts of string for just intonation and for Woolhouse's temperament are as follows:

	Notes	G.	A.	B.	C.	D.	E.	F♯.	G.	
Lengths	Just	13	11·56	10·40	9·75	8·67	7·80	6·93	6·50	inches
	Woolhouse . . .	13	11·65	10·44	9·71	8·70	7·80	6·99	6·50	„
	Differences . . .	0	0·09 flat	0·04 flat	0·04 sharp	0·03 flat	0	0·06 flat	0	„

(As may be seen from the foregoing table of temperaments, violin players already depart more than this when adapting themselves to the equal temperament.)

(vi.) *Chief Intervals within an Octave.*—In the accompanying Table II. are exhibited the chief intervals within an octave. The first column gives the names of the intervals, the other four columns give their values classified under the various systems of tuning, viz. just intonation, mean-tone, Woolhouse's Cycle of Nineteen, and equal temperament. In the column for just intonation the values of the intervals are expressed both in *cents* (or *hundredths* of the *semitone* of equal temperament) and in the ratio of frequencies of the higher to the lower tone of the interval. Thus, for the major third, between the notes C and E, we have the interval expressed as 386 cents and also as 5 : 4. In the other columns the ratios are not given, as they would usually be cumbrous and of very little value. In the column for mean-tone, after the values of the intervals in cents, are inserted in round brackets the values for Huygens's Cycle of Thirty-one, which are very nearly the same. The reader may be here reminded that the various intervals in Bosanquet's Cycle of Fifty-three are practically indistinguishable from those of just intonation; they are accordingly omitted from the table entirely.

§ (7) Historical Pitches.—Widely different pitches have been adopted for the scales at various times and places and for distinct musical uses. A few typical pitches, ranging over five centuries, are given in chronological order in the accompanying Table III. It will be seen from this that the church and chamber pitch began very high, fell, rose, and fell again. The representative European pitches remained with but small fluctuations for more than a century. The orchestral pitch, after remaining fairly steady for some time, rose to a climax, but a tendency to revert to a lower value showed itself in 1896, and this low or flat pitch is still often used. The pitches are specified by the frequency of the note A in the treble staff and also by the intervals in cents above an imaginary lowest pitch. It will be seen from the column of cents that the whole range of church pitches in use covers seven semitones of the equal tempered scale! The old European pitches, on the other hand, varied less than a single semitone. The same remark applies to the modern orchestral pitches, but the lowest of these latter is above the highest of the former.

II. Propagation of Sound

§ (8) Source of Sound.—Sounds have their origin in a rapid to-and-fro motion of some body, solid or fluid, which is then said to be in vibration. This vibrating body may be the string of a harp, piano, or violin, the column of air in an organ-pipe, flute, clarinet, or trumpet, the skin of a drum, the prong of a tuning-fork, or the sides of a church bell. In any of these cases the rapid to-and-fro motion thus started may spread by a wave-motion in the atmosphere somewhat as ripples spread on the surface of still water on a pond round the spot where a stone fell in. But though there is a general resemblance between the two cases, some contrasts need to be noticed. Thus, on the surface of water the spreading occurs in all horizontal directions from the central source of disturbance. Hence the ripples are *circles*. In the atmosphere (if unobstructed) the spreading is in all directions,

up and down as well as horizontal. In consequence, we have a *spherical* sound wave spreading round its central source. Again, in the case of ripples on water, though the spreading is horizontal the to-and-fro motion of the water is *partly up and down.* But, in the case of sound in the air, while the waves of sound spread by advance along any and every straight line from the origin, the to-and-fro motion of the air at any point occurs *wholly in the line* along which the wave is there advancing. For this reason sound waves are said to be *longitudinal.* Thus, while the ripples in water show crests and troughs the sound waves in air have places of compression and rarefaction. Yet a third contrast may be referred to here. The ripples are started by the stone itself; the sound waves often need the intervention of something else between them and the origin. Thus the string of a violin needs the bridge and belly of the instrument, otherwise no appreciable sound would be heard.

TABLE II

CHIEF INTERVALS WITHIN AN OCTAVE

Names of Intervals.	Values of Intervals in various Systems.			
	Just Intonation.	Mean-tone.	Woolhouse's Cycle of 19.	Equal Temperament.
Ellis's cent . . .	1/1200 of octave	..	..	One hundredth of semitone of which 12 = an octave
Skhisma (2 cents nearly)	Excess of 8 fifths and 1 major third over 5 octaves = 1·953 cents	..	..	Practically = 1/100 tone
Comma of Didymus (22 cents nearly)	Excess of 4 fifths over 2 octaves and 1 major third = 21·506 cents or 81 : 80			
Pythagorean comma (23 cents nearly)	Excess of 12 fifths over 7 octaves = 23·46 cents			
Step or smallest interval of cycle	(For Bosanquet's cycle of 53 to octave = 22·6415 cents)	(For Huygens's cycle of 31 to octave = 38·71 cents)	Cents 63·157	Cents 100
	Cents	Cents		
Chromatic semitone .	{D to D♯ 71 or 25 : 24) / C to C♯ 92 or 135 : 128}	76·1 (77·4*)	(say) 63·16	100
Diatonic semitone .	112	117·1 (116·1)	126·31	100
Whole tone. . .	{D to E 182 or 10 : 9 / C to D 204 or 9 : 8}	193·2 (193·5)	189·47	200
Minor third . .	C to E♭ 316 or 6 : 5	310·3 (309·68)	315·79	300
Major third . .	C to E 386 or 5 : 4	386·3 (387·1)	378·94	400
Fourth	C to F 498 or 4 : 3	503·4 (503·23)	505·26	500
Tritone, F to B . .	C to F♯ 590 or 45 : 32	579·5 (580·64)	568·37	600
Fifth	C to G 702 or 3 : 2	696·6 (696·77)	694·73	700
Minor sixth . .	C to A♭ 814 or 8 : 5	813·7 (812·90)	757·88	800
Major sixth . .	C to A 884 or 5 : 3	889·7 (890·32)	884·21	900
Harmonic seventh / Minor seventh (Fourth + fourth) / Acute minor seventh (Fifth + minor third)	C to ⁷B♭ 969 or 7 : 4 / C to B♭ 996 or 16 : 9 / C to B♭′ 1018 or 9 : 5	1006·8 (1006·45)	1010·53	1000
Major seventh . .	C to B 1088 or 15 : 8	1082·9 (1083·87)	1073·69	1100
Octave	C to c 1200 or 2 : 1	1200	1200	1200

* Numbers in ( ) refer to Huygens's cycle of 31.

Sounds may also advance by a wave motion through any other gases, through liquids (as the water of lakes or the open ocean), and through solids (as the crust of the earth itself). The speeds of sound in these different substances will be different and will also vary slightly with the state of each substance.

For mathematical and physical treatment of these cases see E. H. Barton's text-book on *Sound* (London, 1922); J. W. Capstick's

Sound (Camb., 1913); or Alex. Wood's *Physical Basis of Music* (Camb., 1913). A general account of the more important results is given in the following articles.

§ (9) Theoretical Speeds of Sound in Gases.—It can be shown mathematically that the speed of a wave motion in any suitable elastic medium is the square root of the quotient—*elasticity concerned* divided by *density*. Thus, if v is the speed of sound in a gas whose appropriate elasticity is E and density D, we have the relation [1]

$$v=\sqrt{\left(\frac{E}{D}\right)}. \quad . \quad . \quad . \quad (1)$$

Further, in the case of sound in gases the elasticity in question is known to be its pressure (P) multiplied by a factor (k) greater than unity. This factor (k) is the ratio of the two specific heats of the gas under constant pressure and at constant volume respectively. Its exact value depends upon the number of atoms in the molecules of the gas. It approaches 5/3 for a monatomic gas or vapour (*e.g.* argon or mercury vapour), it is about $\sqrt{2}$ for a gas with diatomic molecules, and is about 5/4 for gases with triatomic molecules. Thus for air $k=1{\cdot}408$ and for steam $1{\cdot}26$ nearly.

Thus, if we now write U for the volume of the gas per unit mass (this quantity being the reciprocal of the density), equation (1) becomes

$$v=\sqrt{(kPU)}. \quad . \quad . \quad . \quad (2)$$

Table III

Typical Pitches in Chronological Order

Date.	Frequency of Treble A. Church and Chamber.	European.	Orchestral.	Cents above Lowest.	Place.	Instrument or Authority.
..	..	370	..	0	..	Imaginary lowest pitch to reckon from
1361	505·8	..	..	541	Halberstadt	Organ
1511	377	..	..	33	Heidelberg	Arnold Schlick's low pitch
1619	567·3	..	..	740	North German	Church pitch
1636	563·1	..	..	726	Paris	Mersenne's chamber pitch
1640	457·6	..	..	368	Vienna	Organ
1648	373·7	..	..	17	Paris	
	402·9	..	..	148	Paris	Spinet
1666	..	..	437·1	289	Worcester	Cathedral organ (by Thomas and Renatus Harris)
1700	384·3	..	..	66	Lille	Old fork
1751	..	422·5	..	230	England	Handel's fork
1754	..	415	..	199	Dresden	Organ
1776	414·4	..	..	196	Breslau	Clavichord
1788	..	427·8	..	251	Windsor	St. George's Chapel organ (measured by Ellis, Feb. 1880, while still in mean-tone)
1810	..	430	..	260	Paris	Fork
1820	..	422·5	..	230	Westminster	Abbey organ
1824	..	425·8	..	243	Paris	Opera pitch
1828	..	433·2	..	273	London	Sir G. Smart's own philharmonic fork
1854	398·7	..	..	129	Lille	Old organ
1856	..	..	445·8	323	Paris	Opera
1859	..	..	437	288	Toulouse	Conservatoire
1877	..	422·5	..	230	England	Curwen's "Tonic Sol-Fa," C=507
		..	444·6	318	London	St. Paul's Organ
1878	..	436·1	..	285	London	Fork to which organ tuned at H.M. Theatre
		..	441·2	305	London	Covent Garden Opera
		..	452·9	350	London	Kneller Hall Military School (high pitch or *old* philharmonic) B♭=479·3 at 60° F.
1879	..	..	455·3	359	London	Erard's concert pitch
1880	..	..	460·8	380	U.S.A.	Highest New York pitch
1885	..	..	452	346	London	International Exhibition of Invention and Music
1896	..	..	439	297	London	*Low* pitch or *new* philharmonic B♭ 465, C 522; all at 60° F.

[1] See "Vibrations of Air."

In using these equations to obtain numerical values for any gas in a specified state, the units used must be in agreement and on some definite system. Thus, if the speed is required in centimetres per second, the pressure P must be in dynes [1] per square centimetre and the density D in grams per cubic centimetre (or its reciprocal U in cubic centimetres per gram). If the speed is required in feet per second, the change could be effected by dividing the previous result by 30·48 (the number of centimetres in a foot); or, one might begin by putting the pressure in poundals [2] per square foot and the density in lbs. per cubic foot and so get the result direct in feet per second. The ratio k is a *pure number* and has no units. It remains, therefore, the same whatever system of units is in use.

Example.—To find, in cm. per second, the speed of sound in dry air at 0° C. at a barometric pressure due to 76 cm. of mercury (of density 13·6), the density of the air then being 0·00129 gm. per c.c. and the value of k being 1·408. Then we have

$$v=\sqrt{(1{\cdot}408\times 76\times 13{\cdot}6\times 981\div 0{\cdot}00129)}$$
$$=33{,}265 \text{ cm. per sec. nearly.} \quad (2a)$$

Effect of Moisture.—The ratio of the specific heats in water vapour is about 1·26 and its density is 5/8 that of dry air at the same pressure and temperature. Hence kP/D for water vapour has a value 1·43 times that for air at same pressure and temperature. The speed of sound in air is accordingly increased by the presence of moisture. This increase depends on the amount of water vapour present, and may amount to about two-thirds of a metre per second.

§ (10) Speed affected by Temperature but not usually by Pressure.—By Boyle's law the product of pressure and constant volume of a given mass of gas is constant for temperature. Hence we see from the form of equation (2) that the speed of sound in a gas is *constant if the temperature be constant, whatever the pressure may be* within ordinary limits.[3] For if P increases U correspondingly decreases so as to keep the product PU constant. But if the temperature rises, then the value of PU increases by $\frac{1}{273}$ of its value at 0° C. for every 1° C. rise of temperature. Thus, if v is the speed of sound at t° C., v_0 that at 0° C., and $(\text{PU})_0$ is the value of the product at 0° C., we may rewrite (2) in the form

$$v^2=k(\text{PU})_0\left(1+\frac{t}{273}\right)=v_0{}^2\left(1+\frac{t}{273}\right). \quad (3)$$

[1] At the sea-level in London, one gram weighs about 981 *dynes*.

[2] At the sea-level in London, one pound weighs about 32·2 *poundals*.

[3] Koch in 1907 showed that in air at 0°, but under 25 atmospheres pressure, the speed of sound was 1·008 times that in air under atmospheric pressure.

For moderate changes of temperature this becomes

$$v=v_0\left(1+\frac{t}{546}\right) \text{ nearly.} \quad (4)$$

Thus for sound in air, if we take 33,200 cm./sec. for v_0, we have

$$v=(33{,}200+61t) \text{ cm./sec. nearly.} \quad (5)$$

The equivalent approximate formula in feet per second is

$$v=(1089+2t) \text{ ft./sec. nearly,} \quad (6)$$

if the temperature be measured in degrees Centigrade.

§ (11) Theoretical Speeds of Sound in Liquids.—In the case of liquids the wave motion of sound is still a longitudinal one, as in the case of gases. The elasticity called into play is also, as in gases, the volume elasticity or bulk modulus. But it is no longer easily obtained from the pressure, as for gases. On the contrary, the elasticity for a liquid must be found for each case from the experiments on it that have been made by recognised workers in this domain. Let the value of this elasticity be denoted by K for a given state of some liquid, and let M denote its reciprocal called the compressibility, then we have for the speed of sounds in the liquid the equation

$$v=\sqrt{\left(\frac{\text{K}}{\text{D}}\right)}=\sqrt{\left(\frac{1}{\text{MD}}\right)}, \quad (7)$$

where D is the density of the liquid.

Often, in tables on the subject, the *compressibility* of liquids is given. Thus for water at 1-25 atmospheres pressure and 15° C. Amagat found the compressibility to be $48{\cdot}9\times 10^{-12}$ for an increase of pressure of one dyne per sq. cm. The density of water at 15° C. is 0·99912 gm. per c.c. Thus the value of the speed of sound in it would be given by

$$v_{15}=\sqrt{\left(\frac{10^{12}}{48{\cdot}9\times 0{\cdot}99912}\right)}$$
$$=143{,}166 \text{ cm./sec. nearly.} \quad (7a)$$

If the density be taken at 8° C. as 0·9999 gm. per c.c. and the elasticity be supposed to remain practically constant, we should have

$$v=\sqrt{\left(\frac{10^{12}}{48{\cdot}9\times 0{\cdot}9999}\right)}$$
$$=143{,}000 \text{ cm./sec. nearly.} \quad (7b)$$

It may be seen that although the much greater density of water than air tends to lessen the speed of sound in it, the still greater disparity of their elasticities overcomes that drawback. Thus the speed of sound in water is over four times that in air.

§ (12) THEORETICAL SPEED OF SOUND IN SOLIDS.—Suppose first that the solid is only a *rod*, and therefore quite free to shrink laterally where stretched and to bulge where compressed lengthwise. Then if Y denotes the value of Young's modulus [1] for the material of the rod and D its density, we have the following expression for the speed of sound in it:

For a solid rod,

$$v=\sqrt{\left(\frac{Y}{D}\right)}. \qquad (8)$$

As an example consider brass, for which $Y=10^{12}$ dynes per sq. cm. and $D=8\cdot4$ gm. per c.c. nearly. Then

$$v=\sqrt{(10^{12}\div8\cdot4)}$$
$$=345{,}033 \text{ cm./sec. nearly,} \qquad (8a)$$

or over *ten* times the speed in air.

For an extended solid like the crust of the earth, where there is no freedom for bulging sideways at places compressed endwise, a quite different relation holds. In this case let K be the bulk modulus or volume elasticity and N the rigidity (or shape elasticity). Then the expression for the speed of sound takes the form:

For an extended solid,

$$v=\sqrt{\left(\frac{K+\frac{4}{3}N}{D}\right)}. \qquad (9)$$

Suppose an extended mass of brass existed, and let it be required to find the speed of sound in it. Then, taking the value of K to be $10\cdot65\times10^{11}$ and N to be $3\cdot5\times10^{11}$, each in dynes per sq. cm., we obtain

$$v=\sqrt{\{(10\cdot65+\tfrac{14}{3})10^{11}\div8\cdot4\}}$$
$$=427{,}061 \text{ cm./sec. nearly.} \qquad (9a)$$

It may be noticed that the value of the speed for the extended solid is in this case nearly 5/4 of that in a rod of the same material.

§ (13) EXPERIMENTAL DETERMINATIONS OF SPEED OF SOUND IN AIR. PARIS, 1738.—The earliest exact determination of the speed of sound in the open air seems to have been made in 1738 by members of the Paris Academy. The stations used were the observatory at Paris, Montmartre, Fontenay-aux-Roses, and Montlhéry, involving a total distance of about 18 miles. The experiments were made at night by firing cannons alternately at the two end stations. The calculations depended upon seeing the flash and hearing the report at the other stations. Then the distance in question divided by the time elapsed between the receipts of the flash and the report gave the first rough value for the speed sought.

The time required by light to cover this distance is justly neglected, since the speed of light is about 186,000 miles per second. The experiments were continued for some time under different conditions of the atmosphere as to wind, temperature, barometric and hygrometric states.

The conclusions arrived at from these investigations may be stated as follows:

(i.) The speed of sound is independent of air pressure.

(ii.) It increases with the temperature of the air.

(iii.) It appeared to be the same at each distance from the source.

(iv.) The speed of sound with the wind is the sum, and against it is the difference, of the speeds of sound and of the wind.

(v.) The speed of sound in still dry air at 0° C. (deduced by Le Roux from above experiments) is 332 metres per second.

Four of the above conclusions have been confirmed by other workers at various times and places and under very diverse conditions.

§ (14) SPEED OF SOUND IN AIR INCREASES WITH INTENSITY.—The one conclusion that cannot be accepted to-day for ordinary conditions is the third. This would make it appear that the speed of sound is independent of the intensity. It might not have varied appreciably in the experiments in question so far as the stations used permitted an analysis of the matter. But it is now known that with sufficiently great intensities the speed is appreciably increased. This was shown in 1864 by Regnault at Versailles by experiments in the open air, two distances, 1280 and 2445 metres, being used, and reciprocal firing of guns adopted. The instants of firing and of receipt of the sound were electrically recorded, and the following results were obtained:

For the distance of 1280 metres, $v_0=331\cdot37$ m. per sec.

For that of 2445 metres, $v_0=330\cdot7$ m. per sec.

Hence by these experiments it is shown that nearer the gun, where the sound is more intense, the speed is slightly greater. This is in accord with theory. For when the sounds are sufficiently intense the temperature in the compressions is appreciably raised, and the speed increased accordingly

This increase of speed with intensity of the sound is well shown by the photographs of bullets in flight, due to Professor C. V. Boys, taken in London in 1892 (see *Fig.* 2). If a projectile moves much quicker than the normal speed of sound, the air in front of its nose is highly compressed and so heated until the appropriate speed of sound is that of the bullet. So then the bullet has in front of it and round its sides a shell of compressed air like the breast wave round a swan when

[1] This is the quotient of stretching force per unit area divided by elongation per unit length. See also "Vibrations of a Rod."

advancing in still water. In each case, the quicker the motion of the object through the medium the sharper will be the advancing nose of the wave. Further, in each case, the relation of the speed of the moving object

FIG. 2.—Boys's Photograph of Bullet in Flight.

to that of the normal waves in the medium can be ascertained from the geometry of the breast wave.

§ (15) EXPERIMENTAL DETERMINATION OF SPEED OF SOUND IN WATER.—The classic experiments on the speed of sound in water are those by Colladon and Sturm carried out in 1826 in the Lake of Geneva. The stations were two boats moored at a measured distance apart. From one a bell hung in the lake and was sounded by a lever whose upper end by the same motion fired some gunpowder, thus signalling to the other boat the instant when the sound started. The sound travelling through the water was received at the other boat by a funnel whose mouth was below in the water and upper end applied to the observer's ear. The mean temperature was estimated as 8°·1 C., and the speed found was 1435 metres per second.

It may be noticed that the theoretical result calculated earlier is in fair agreement with this, especially when it is considered that the water in the open lake must be very different from that experimented upon in the laboratory for its compressibility.

§ (16) SPEED OF SOUND IN WATER INCREASES WITH INTENSITY.—Threlfall and Adair published in 1889 their experiments on the speeds of sound-waves from explosions under water in Port Jackson harbour, Australia. Charges of gun-cotton were used; the firing was electrical and gave a signal on a chronograph, on which also was recorded the instant when the sound reached an india-rubber diaphragm immersed at the far station 150 metres or more distant. The normal speed for feeble sounds in the water was calculated to be about 1500 metres per second; the observed speed of the explosion wave from 9 oz. of gun-cotton was 1732 m./sec., and from 64 oz. of gun-cotton was 2013 m./sec.

§ (17) EXPERIMENTAL DETERMINATION OF SPEED OF SOUND IN IRON.—For this purpose Biot used 376 cast-iron pipes with a total length exceeding 950 metres. A bell at one end was struck and its sound travelled through the iron walls of the pipes as well as through the air inside them. At the far end the sound through the iron was received first and that through the air arrived at a measurable time, t seconds, later. Hence, if the total length of the pipes is L cm., the speed of sound in cast-iron u cm./sec., and that in the air inside the pipes v cm./sec., we have for the times of travel through air and through iron L/v and L/u respectively. But their difference is t. Thus

$$\frac{L}{v}-\frac{L}{u}=t,$$

whence

$$u=\frac{Lv}{L-vt}.$$

The value thus found for cast-iron was about 350,000 cm./sec. or 3500 metres/sec.

§ (18) REGNAULT'S DETERMINATION WITH ROUGH PIPES.—This determination was the result of an elaborate set of experiments carried out with the water-pipes newly laid in Paris in 1862–63. A pistol, explosions, and musical instruments were used as sources of sound, the method of measuring time being electrical. The shot broke that part of an electrical circuit just in front of the pistol, thus causing a shift of the mark traced on a smoked rotating drum. This shift registered the start of the sound on its course along the air inside the pipes. The sound on reaching the far end of the pipe moved a very fine membrane. This made an electric contact and so completed a circuit, thus causing the reversed shift of the mark on the drum. Side by side with this mark were (1) a set of dots indicating seconds given by a pendulum, and (2) a wavy trace due to a tuning-fork. Thus the time elapsing between the start and arrival of the sound (corresponding to the length on the mark between the first and final shifts) was measured in seconds by the dots

and in thousandths of a second by the wavy trace. Regnault tried to allow for the time lag of the membrane, which was greater for feebler sounds. The length of the pipe divided by the correct time gave a quotient which represented the speed of the sound under the experimental conditions. The pistol was inserted at the beginning of the pipe, which was there closed by a disc, and the far end was closed by the membrane. Thus the sound could pass to and fro several times, suffering partial reflections at the ends. Pipes of three diameters were used and a number of lengths tested. The variations of speed with length and diameter are shown in the accompanying graphs of *Fig.* 3.

From these curves it may be seen (1) that the speed of sound tends to a lower limit when very feeble owing to the great distance traversed,

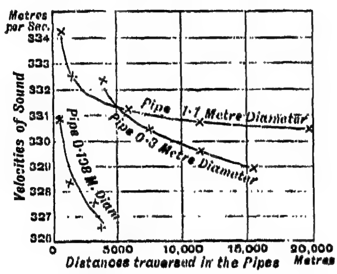


Fig. 3.—Regnault's Speeds in Pipes.

and (2) that this limiting speed is greater in wide pipes than in narrow ones. Regnault considered that in the pipe of 1·1 metre diameter its sides were practically without effect. He therefore gave 330·6 metres per second as the speed of sound in air at 0° C. in an infinitely wide pipe (or, by taking the sound in its feeblest state, 330·3 m./sec.). The humidity of the air in these experiments on pipes was observed and corrected for. Rink by a special analysis of Regnault's results deduced the value 330·5 m./sec. for the speed of sound through air in a pipe 1·1 metre diameter.

In other experiments Regnault failed to detect any difference in the speed of sound, although the pressure was increased to more than five times its first value (viz. from 247 to 1267 mm. of mercury).

Violle and Vautier, in 1905, by experiments on pipes, found that within the range of frequencies from about 32 to 640 vibrations per second the speed of sound in air does not change by one-tenth per cent.

§ (19) Blaikley's Determination with Smooth Pipes.—This account of researches in the speed of sound through air in smooth brass tubes was published by D. J. Blaikley in 1883–84. He felt that in any open-air experiments the results were usually vitiated by uncertainties as to mean temperature and hygrometric state of the air over the range in use. Further, he noticed that in Regnault's experiments with pipes the decrease of speed there found would, if extended to the diameters used in brass instruments (in which Blaikley was specially interested), lead to smaller values than those actually experienced. This discrepancy is attributed to the roughness of Regnault's pipes. Blaikley also feared that the membrane used by Regnault might lead him to an underestimation of the speed, and this surmise was then supported by special experiments. Dulong, in 1829, had estimated the speed of sound by experiments on organ-pipes specially chosen as giving a tone of good musical quality. To this Blaikley objected, because such a tone consists of the prime (or fundamental) accompanied by a retinue of fainter overtones of frequencies double, treble, quadruple, etc. that of the prime. Now these overtones, when elicited *separately*, have frequencies slightly different in value from the exact double, treble, etc. of the prime frequency. Then, on blowing the pipe, these naturally inharmonious overtones are forced into the strict relation to the prime expressed by the ratio numbers 2, 3, 4, etc. But this forcing into tune of the overtones and their prime does not operate solely by changing the overtones, but also may *slightly disturb the prime itself* in arriving at the requisite adjustment. The prime tone is in consequence slightly different in pitch from that which, without such constraint, corresponds to the true speed of sound and the wave-length under consideration. Hence, from the length of the pipe and this constrained pitch a vitiated speed of sound would be inferred.

In view of the above criticisms Blaikley preferred to experiment with a special form of pipe which resembled the glass chimney of an oil lamp instead of being parallel throughout as is an ordinary organ-pipe. That is to say, the pipe used has a bulb or pear-shaped portion in the first quarter wave-length near the mouth. Beyond the bulb there was a considerable length of parallel cylindrical pipe in which worked a sliding plug. This pipe, when blown, gave the fundamental or prime tone alone. Consequently its pitch was not disturbed by being blown into tune with any overtones, as these were entirely absent. The blast was obtained from a fan, the wind from which passed a regulating bellows. The bellows pressure was 2·5 inches of water column, and that in the mouth less than 0·1 inch of water, much under the lowest which Regnault had found to alter the speed. Both the temperature and humidity of the air were observed. The pitch was taken from

a Koenig fork of 256 vibrations per second and the pipes were set to give a beating rate of about four per second, the lengths being gauged by standard rods and a micrometer.

Experiments were made with smooth brass tubes of five sizes, the frequencies of the tones ranging from about 131 to 323 per second. The results are shown in Table IV.

Blaikley's values show that the diminution of speed is proportional to the reciprocal of the diameter of the pipe. They also favour the view that it is proportional to the reciprocal of the square root of the frequency of the tone sounded in the pipe.

Blaikley applied to the above mean values obtained in Paris (*a*) in the open air and (*b*) as deduced from pipes.

§ (20) Hebb's Telephone Method for Speed of Sound.—In 1904 T. C. Hebb made a very careful determination of the speed of sound in air by use of parabolic reflectors and telephones. No pipes were used, as they were regarded as introducing complications. All long-distance methods were objected to on the following grounds:

(*a*) Very loud sounds must be used, so possibly involving a different speed near the source.

(*b*) It is almost impossible to correct accurately for wind, temperature, and humidity over long ranges.

Table IV

Blaikley's Speeds of Sound for Dry Air at 0° C. in Brass Tubes

Diameters of Tubes.	11·43 mm.	19·05 mm.	31·71 mm.	52·91 mm.	88·19 mm.
Speeds of sound from separate experiments in metres per second.	324·533	327·09	328·72	329·90	330·20
	324·234	327·14	328·74	329·84	330·46
	..	326·98	328·78	329·84	330·02
	..	326·70	328·72	329·70	329·72
	..	327·09	328·72	329·95	329·99
	..	326·69	328·89	329·80	330·41
	..	326·99	328·76	329·53	330·09
	..	326·79	328·84	329·56	330·06
	..	326·70	328·84	329·65	330·10
	..	326·85	328·83	329·48	330·20
Mean speeds for each diameter . . .	m./sec. 324·383	m./sec. 326·90	m./sec. 328·78	m./sec. 329·72	m./sec. 330·134
(Probable errors of mean speeds, added by writer in 1921).	..	m./sec. $\pm 0{\cdot}037$	m./sec. $\pm 0{\cdot}013$	m./sec. $\pm 0{\cdot}035$	m./sec. $\pm 0{\cdot}046$

(The probable errors were not given by Blaikley himself, but have been added for the present work to show how accurate this investigation is and how closely it may be trusted. If the probable error be both added to and taken from the mean value to which it applies, limits will be obtained between which the true value is as likely to lie as to be outside of them.)

of the speed for each sized pipe a modification of Helmholtz's formula in order to calculate the speed of sound in free air. He thus deduced the mean value

$$v_0 = 331{\cdot}676 \text{ metres per second}$$

as his final determination for the speed of sounds of feeble intensity in dry open air at 0° C. (To this we might now add a probable error of $\pm 0{\cdot}04$ m./sec.)

The corresponding value for the ratio of specific heats for dry air was $k = 1{\cdot}4036$.

For purposes of application to brass instruments this investigation must be regarded as the standard of appeal, since it was obtained by direct use of such tubes. Further, the speed deduced for the open air, though obtained indirectly, must be regarded as of extremely high value. It is seen to lie between the two values already quoted, as

(*c*) The *lag* (or *personal equation*), either of a human operator or of some recording device, is involved.

In accordance with these principles a novel method was adopted by which it was hoped all the above drawbacks could be obviated. The experiments were carried out in a room 120 feet long, with a whistle as the source of sound and blown so steadily as to maintain its frequency to an accuracy of 1 in 5000. This source was placed at the focus of a parabolic mirror made of plaster of Paris. From the mirror proceeded a parallel beam of sound waves along the room to a second similar mirror which converged them to its focus. At the source of sound was a telephone transmitter connected to a battery, and to *one of two primaries* of a special induction coil. At the focus of the second mirror was another similar telephone transmitter, also

connected to a battery and to the *other primary* of the induction coil. The *secondary* of this induction coil was connected to a telephone receiver. Suppose the whistle to be sounding and the mirrors to be such a distance apart that the sound heard from the telephone receiver is a maximum. Then the effects of the two primaries of the induction coil must be in such relation as to make their combination most powerful, *i.e.* their vibrations are in the same phase and so are mutually helpful. Now let the distance between the mirrors be changed till the sound from the telephone receiver is a minimum. Then the primaries must be just opposite in their effects now. If the separate effects, which could thus be made additive or subtractive for different distances, were practically equal, then the minima effects would be reduced almost to zero, and the corresponding positions could be accurately observed. Let the frequency of the whistle be N per second and its period, or time of one complete vibration, be T second. Then, if the speed of sound in air is v and the distance in air passed over by sound in one period is L cm., we may write

$$v = \frac{L}{T} = NL.$$

The length L is called the *wave-length* in air of the sound used. And it may be seen that it is on changing the distance between the mirrors by *half the wave-length* that we change the effects of the two primaries from the complete additive to the complete subtractive relation, and so change the result heard from maximum to minimum. Hence, by locating a number of successive maxima and minima the wave-length L was found. In this determination an accuracy of about one in a thousand was reached.

The determination of the pitch of the whistle was made by tuning it in unison with a fork which was itself compared with a 512 fork by wavy traces on a smoked disc. The 512 fork was compared in the same way with a pendulum, and then the pendulum with a clock.

Thus knowing L and N, their product gave v the speed of sound sought. The final determination for the speed of sound in dry air at 0° C. was

$$v_0 = 331.29 \text{ metres per second,}$$

with a probable error of ± 0·04 m./sec.

§ (21) Speeds of Sound in Various Media.—The foregoing, together with other determinations of the speeds of sound in

Table V

Speeds of Sound in Various Media

Medium.	Condition.	Speed.	Determined by.
Gases.		Metres per second.	
Open air	Still and dry at 0° C.	327	Paris Academy in 1738
„	„ „	332	Paris Academy and Le Roux
„	„ „	331·2	Bureau des Longitudes, 1822
„ over 1280 metres	„ „	331·37	Regnault, 1864, *i.e.* intense sounds slightly faster
„ 2445 metres	„ „	330·7	
Air in rough pipes 1·1 metre diameter	„ „	330·6-330·3	Regnault, 1862-63
Air in smooth pipes 88 mm. diameter	„ „	330·13	Blaikley, 1883-84
Open air (inferred)	„ „	331·676	
Open air (inferred)	„ „	331·1	Violle and Vautier, 1888
Air in laboratory	„ „	331·29 ± 0·04	Hebb, 1904
Dry air free from CO_2	„ „	331·92 ± 0·05	Thiesen, 1908
Dry air	„ „	330·63	Esclangon, 1918
Liquids.			
Fresh water (Lake Geneva)	8° C.	1435	Colladon and Sturm, 1826
Sea water (explosions from 9 oz. gun-cotton)	18° C.	1732	Threlfall and Adair, 1889
Sea water (explosions from 64 oz. gun-cotton)	18° C.	2013	
Sea water	14·5° C.	1503·5	Marti, 1919
Solids.			
Brass in bulk	..	4270 (say)	(Approximate)
Iron (wall of cast pipes)	..	3500	Biot
Iron (in bulk)	..	4300 (say)	(Approximate)

various media, are collected in the accompanying Table V. for ready reference.

§ (22) Architectural Acoustics. (i.) *Sabine's Work.*—It is well known that some churches, halls, theatres, or other buildings are acoustically unsatisfactory. They prove trying to the speakers or musical performers, to the audience, or to both. Others are good for speech but bad for music, or *vice versa*, while yet others may be acceptable for both speech and music. The principles and problems underlying the successful design or remedial treatment of any auditorium to be used for a specific purpose for long failed to receive the attention which, from the importance of the subject, they deserved; but considerable progress in this direction has been made during the present century.

The sounds uttered by a speaker or musical instrument reach the auditor somewhat enfeebled by the distance passed over, but have, on the other hand, various echoes, reverberations or distortions added to them by the floor, walls, ceiling, furniture, and occupants of the room. Thus two questions accordingly present themselves for quantitative treatment: (*a*) What, if any, enfeeblement or additions are desirable in the case of speech, song, and instrumental music? (*b*) How may those desired effects be realised with approximately quantitative accuracy? The subject has been investigated along both these lines by the late W. C. Sabine, who states:[1] "In order that hearing may be good in any auditorium it is necessary that the sound should be *sufficiently loud*; that the simultaneous components of a complex sound should maintain their *relative intensities*, and that the *successive sounds* in rapidly moving articulation, either of speech or of music, should be *clear and distinct*, free from each other and from extraneous noises." These three are the necessary and sufficient conditions for good hearing. Scientifically, the problem involves the absence (*a*) of *prejudicial enfeeblement*, (*b*) of *distortion* of quality of sounds, and (*c*) of too *prolonged reverberation*. From the constructive point of view the problem involves the *size* and *shape* of the auditorium and the nature of the *materials* (including the audience itself) whose surfaces are exposed to the sounds.

(*Ibid.*) "Sound, being energy, once produced in a confined space, will continue until it is either transmitted by the boundary walls or is transformed into some other kind of energy, generally heat. This process of decay is called absorption. Thus, in the lecture-room of Harvard University, in which this investigation was begun, the rate of absorption was so small that a word spoken in an ordinary tone of voice was audible for 5·5 seconds afterwards. During this time even a very deliberate speaker would have uttered twelve or fifteen succeeding syllables. Thus the successive enunciations blended into a loud sound, through which and above which it was necessary to hear and distinguish the orderly progression of the speech. Across the room this could not be done, nor even near the speaker, except with an effort, wearisome in the extreme if long maintained. With an audience filling the room the conditions were not so bad but still intolerable."

"This may be regarded as a process of multiple reflection from walls, ceiling, and floor, losing a little at each reflection until ultimately inaudible. This phenomenon will be called reverberation, including as a special case the echo. In general, reverberation results in a mass of sound filling the whole room and incapable of analysis into its distinct reflections. It is thus difficult to recognise and locate. In the general case of reverberation we are concerned only with the rate of decay of the sound. This rate of decay was gauged by measuring what is inversely proportional to it the duration of audibility of the reverberation or residual sound.

"Broadly considered, there are two, and only two, variables in a room shape (including size), and materials (including furnishings). In designing an auditorium an architect can give consideration to both; in repair work for bad acoustic conditions it is generally impracticable to change the shape, and only variations in materials and furnishings are allowable."

To test the effects of absorption Sabine tested a lecture-room at Harvard empty, without and with cushions. With an organ-pipe as source of sound and a chronometer for timing the duration of audibility after the sound had ceased, he found a time of 5·62 seconds in the absence of cushions. With 8·2 metres of cushions this time fell to 5·33 seconds, and with 17 metres of cushions it fell to 4·94 seconds. With all the seats (436) covered with cushions the time of audibility was only 2·03 seconds. By covering the aisles and platform and placing other cushions on a scaffold till the whole number available (1500) were in use, the duration of audibility fell to 1·14 seconds.

Sabine found that the above duration of audibility of the sound in a room was directly proportional to the total volume of the room, and inversely as the total absorbing power of room empty plus that of audience. Or we may write

$$t = c\frac{V}{R+A},$$

[1] *Sci. Abs.* A, 1915, p. 194.

where t is the time, c is a constant (depending on the units in use), and may be taken as about 0·16 when the metre is the unit of length, or 0·05 when the foot is the unit of length, V the volume of the room, R its absorbing power when empty, A the absorbing power of the audience.

(ii.) *Halls in Paris.*—Marage tested the acoustic properties of six halls in Paris [1] and quoted Sabine's formula as applicable. "In the largest of these halls, the Trocadero, holding 4500, the mean time of resonance was 2 sec. when empty and 1·4 sec. when full. To make himself distinctly heard in this hall a speaker must use a slow utterance, pausing at each phrase. But it was not necessary to use more energy than in addressing 250 in the Physical Theatre of the Sorbonne. In the large theatre of the Sorbonne, holding 3000, the resonance extended almost to 3 sec. empty, but was only 1 sec. or less when full. The acoustic properties of this hall were considered very good."

Marage agreed with Sabine that this time of resonance serves to characterise the acoustic properties of a hall. He found that this duration of the sound varies with the quality, pitch, and intensity of the primary sound; hence a hall good for a speaker may be bad for an orchestra. For a hall to be good acoustically Marage considered that the duration of the resonance should be practically constant at all parts of the hall, for all vowels, and fall between half a second and a second. If the duration of the resonance much exceeds a second, the speaker can make himself understood only by speaking very slowly, articulating distinctly, and avoiding giving to the voice too much energy.

(iii.) *Jaeger's Treatment.*—G. Jaeger carried the treatment of the subject a step further by studying the growth, maximum, and decay of the intensity of sound in a room.[2] He showed that E, the intensity of sound per unit volume in a room, while the source was sustained constant, grew according to the law

$$E=\frac{4P}{vaS}(1-e^{-kt}),$$

where P is the power of the source, v the speed of sound in the air of the room, a the mean absorption coefficient (ratio of absorbed intensity of sound to that falling upon it) of the surfaces in the room (walls, floor, ceiling, audience, etc.), S the total area of these surfaces, $e=2·7183$, $k=vaS/(4$ volumes of room), and t is the time from commencement of sound at source.

Thus the maximum value of E would be $4P/vaS$. The decay of the sound when the source became silent would be given by

$$E=\frac{4P}{vaS}e^{-kt},$$

where t is now the time from cessation of sound at source.

The value of the product aS, occurring in these formulae, is the sum of a number of like products, say $a_1S_1+a_2S_2+a_3S_3$, where a_1, a_2, a_3, etc., are the absorption coefficients for surfaces of areas S_1, S_2, S_3, etc., respectively.

The formulae show that the greater the product aS, the *smaller* is the maximum energy that can be attained in the room by a source of fixed power P, but that the rise to this maximum is *quicker* and the decay *more sudden* also; though these times of rise and fall depend upon the *volume* of the room as well.

(iv.) *Watson's Investigations.*—In 1913 F. R. Watson published an account of a case in which the ventilating arrangements were held to assist the acoustics of the room under notice, namely, the Baltimore Academy of Music.[3] In this hall "the whole supply of fresh air was admitted at the back of the stage, was there warmed, then crossed the stage horizontally *above* the speakers, actors, or musicians, passed through the proscenium, and then, somewhat diagonally toward the roof, across the auditorium in one grand volume and with gentle motion so as almost entirely to prevent the formation of minor air-currents. It was exhausted partially by an outlet in the roof and partly by numerous registers in the ceilings of the galleries. The acoustics of this auditorium were considered excellent. The weakest voice was audible to every seat in the house; sounds such as a sigh, a kiss, or even the simulated breathing of the somnambulist, might be heard in the most distant parts; and all musical effects were exactly appreciated. All singers and speakers agreed in describing the facility with which the voice was used on this stage." It was stated that haphazard currents of air or those intervening between the speaker and the audience are detrimental, and should be avoided.

Later, F. R. Watson described his acoustic treatment of the faulty auditorium at Illinois.[4]

This room was dome-shaped and had many subordinate curved surfaces, which focussed the sounds in an objectionable manner. It seated 2200 and originally showed a reverberation of 9 seconds. After some treatment

[1] *Comptes Rendus*, April 9, 1906, cxlii. 878-880; and *Science Abstracts*, A, 1906, 313.

[2] *Akad. Wiss. Wien. Ber.*, 1911, cxx. 2a, 613-634; *Science Abstracts*, A, 1912, 17.

[3] *Eng. Record*, 1913, lxvii. 265-268; *Science Abstracts*, A, 1913, xvi. 330.

[4] *Univ. of Illinois Bull.*, 1914, xi. No. 73, 3-29; *Science Abstracts*, A, 1914, xvii. 540.

the mean time of reverberation determined from 400 observations was 5·9 seconds, whereas, calculated by Sabine's formula (volume of room 11,800 cubic metres), this time came out as 6·4 seconds. Thick carpets were placed on the stage, a large canvas painting (400 square feet) was introduced, and the glass removed from the skylight in the ceiling. This reduced the time of reverberation of the hall when empty to 4·8 seconds. When quite full the hall showed a still shorter reverberation which was not troublesome. Pronounced echoes, however, still gave trouble. These were prevented by draperies hung in the dome. A sound-"board" was constructed of plaster in a paraboloidal form and acted well for a single speaker, but was objected to by some on account of its appearance. Further, it was useless for a chorus or band, and was accordingly discarded. The auditorium was finally corrected by installing a calculated amount of hair felt on walls that produced echoes. A decorative cloth was applied over the felt to give a pleasing appearance.

Watson afterwards treated the armoury at Illinois.[1] This room presented an unusual case of defective acoustics because of its huge volume and small absorbing power. It was built to fulfil the usual requirements of an armoury in regard to military drills; but, in addition, it had been used occasionally for other large assemblies. The acoustics proved impossible for speaking and music. The room is 400 feet by 212 feet by 93 feet high at the centre of the roof, which is almost semicircular (and thus forms a reminder of the roof of St. Pancras Station, London). Calculated by Sabine's formula the time of reverberation, when empty, was 24 seconds. It was now seen why attempts with a special megaphone, parabolic reflectors, and loud-speaking telephones had all failed.

The final and successful treatment lay in the introduction of canvas curtains to enclose a space 212 feet by 134 feet by 34 feet high. The time of reverberation in this enclosure for an audience of 4500 was then estimated to be 1·1 second. The arrangement was carried out and auditors in all parts of this space could hear and understand the various speakers.

(v.) *General.*—Large, bare, unbroken wall spaces of stone, brick, or plaster may produce objectionable *echoes* and are to be avoided whenever possible in a room intended for speaking or rapid music. Roughnesses or recesses in the walls only count when their size is comparable to the wave-length of the sounds in use. For the quiet speech of a lecturer with a bass voice in a room to hold, say, 100, the wave-length is of the order eight feet. Consequently, in such a case, window recesses two feet wide and six inches deep could scarcely be counted as roughnesses, but alcoves four feet wide and two feet deep might be so reckoned. The same remarks apply to the recessing of the ceilings by beams, coves, mouldings, etc.

If by reason of bare hard walls distinct echoes occur, their objectionable presence may be sometimes reduced by the use of a *sound-board* over the *speaker*. Thus if, by the meeting of the direct and reflected sounds at some one small region in the room, the hearing is there very defective, a *concave* sound-board may focus part of the sound directly to that "dead" place and so improve matters. In other cases it may be advisable to use a *plane* or *convex* sound-board to prevent the chief sound proceeding direct to that surface which produces the objectionable echo, or to cut the sound off entirely from the upper part of the space. Obviously the sound-board is not easily applicable to the case of musicians, so we consider it for the case of a single speaker only and when located at a definite place. Then, since the wave-length may be large, say 8 feet, the sound-board must be correspondingly large or the waves of sound will bend round it and the board will be useless. Again, if the sound-board is required (as usual) to reflect as much as possible, it is better made of plaster on a framework of wood and finished to the exact curvature needed for the purpose in view.

A careful examination of all the circumstances of the case is required before deciding what, if any, sound-board is desirable and might be expected to prove of service.

There are very few, if any, cases in which *wires* strung along or across a building can be expected to be of the slightest use.

It is to be noticed that a room considered to be right for speech may be just a little too *dead* for music. Such a room may be improved therefore by the removal, on musical occasions, of some carpets, curtains, or other deadening trappings which are needed for lectures.

(vi.) *Absorption Coefficients.*—In the accompanying table are given the values of the coefficients of absorption for a number of different materials as determined by W. C. Sabine, H. O. Taylor, and others. These will be found useful to insert in Sabine's formula, either when designing any auditorium, or when the endeavour is to remove the defective acoustics of some room already in use.

The absorption coefficient, a, is defined as the quotient, *intensity of sound absorbed* by a given surface divided by the *intensity of sound falling upon it.*

[1] *Brickbuilder*, 1915, 1-4; *Science Abstracts*, A, 1916, xix. 74-75.

TABLE VI

ABSORPTIONS OF SOUND

Substance.	Coefficient of Absorption a.
Open window	1·000
Audience	0·96 to 0·44
Hair felt (1 inch thick)	0·78 to 0·50
Compressed cork (1½ inch thick)	0·32
Oriental rugs, extra heavy	0·29
Oil paintings and frame	0·28
Asbestos roll fire felt (¾-inch thick)	0·26
Heavy rugs	0·25
Brussels carpet	0·23
Carpet rugs	0·20
Cretonne cloth	0·15
Linoleum, loose on floor	0·12
Pine boards	0·06
Plaster on laths	0·034
Single glass	0·027
Bricks or plaster	0·025
Cheese cloth	0·019

Reference should also be made to the sound-proofing of buildings.[1]

The bulletin on "Sound-proof Partitions" gives a survey of the whole subject with a bibliography of published articles.

§ (23) LONG-RANGE TRANSIT OF SOUND.—When waves of sound are spreading from a powerful source in the open air a number of causes operate to diminish their intensity and may prevent their ever reaching the destination for which they were intended. A number of these cases will be noticed here. Much of our knowledge of such phenomena is due to Tyndall and to the late Lord Rayleigh.[2]

(i.) *Radiation.*—If sound issues from a small source, comparable to a point, and spreads equally in all directions in a still homogeneous atmosphere, then its intensity falls off inversely as the square of the distance from the source. Thus at double the distance the intensity is reduced to a quarter, at treble the distance to one-ninth, and so on.

(ii.) *Temperature Gradient.*—Kelvin has shown that, when undisturbed by winds and sunshine, the air tends to a state which he called *convective equilibrium.* In this state, if air be suddenly taken from one level to another, the expansion or compression consequent on the change of pressure, but *without any gain or loss of heat*, would just bring it to the temperature already possessed by that region. There is thus a lower temperature at greater heights since there the pressure must be less, and the air in ascending to it would suffer expansion and consequent cooling. It can be shown that in such a state, the temperature of the air falls steadily in proportion to the height ascended.

Where this state of things prevails it has been shown by the late Lord Rayleigh that the path of a ray of sound through the air is a catenary with vertex downwards and that if the ray were reversed the same catenary would be described. (A *catenary* is the shape assumed by a uniform and flexible chain or rope fixed at two points and hanging slack at rest under gravity, the vertex being the lowest point.)

(iii.) *Wind Gradient.*—If the wind is everywhere horizontal, and in the same direction, but increases uniformly from the ground upwards, the ray of sound may describe a catenary or something closely like it. But in this case the path is not reversible. Thus, if the direction of the sound is against the wind the catenary has its vertex downwards. But if the sound passes with the wind the catenary described has its vertex upwards. Hence, if cannons are fired at two stations, A and B, it may happen that those fired at A will be plainly heard at B, whereas those fired at B may be inaudible at A.

(iv.) *Regular Reflection.*—Any large solid surfaces, such as a hillside or a cliff, may produce regular reflection as in the case of the ordinary echo. Such reflection may also occur at invisible surfaces in the atmosphere, where its density suddenly changes owing to a change in temperature, or to a change in humidity, or to changes in both. These may be called aerial echoes. Tyndall considered they are probably concerned in the rolling of thunder.

(v.) *Scattering.*—Sometimes the atmosphere, although optically transparent, is partially opaque to sound. It has been found that this is due to the presence of patches or layers of alternately different densities owing to sudden changes in temperature, in humidity, or in both. In this state of things, no regular reflection occurs such as produces an echo. But, on the other hand, the sound is scattered from each such irregular surface, and so the intensity of the direct sound is, in consequence, distinctly diminished. This phenomenon is analogous to the reddening or partial obscuration of the sun when setting in a smoky atmosphere and to the scattering of light from the interior of a turbid liquid or of an opal glass. For in these latter cases the optical media cited, while failing to give a reflection in the regular manner needed to form a mirror image, yet stop and scatter so much of it that the quantity passing directly onwards is appreciably reduced.

This state of semi-opacity in the acoustic sense was carefully investigated by Tyndall, and illustrated by the following lecture

[1] See "Insulation of Sound," W. C. Sabine, *Brick-builder*, 1915, xxiv. 31. "Sound-proof Partitions," F. R. Watson, *Univ. of Illinois Bulletin*, No. 127, 1922.

[2] See Tyndall's *Sound*, London, 1895.

experiment. A number of alternate layers of carbon dioxide and coal gas were interposed between a bell and a sensitive flame. The flame then remained unaffected by the bell, though the region which was acoustically opaque was optically transparent. When the two gases were turned off and air had diffused into the space, the flame responded readily to the bell.

Suppose the atmosphere has many of these patches which scatter the sound, so that its acoustic action may be compared to the optical effects of a turbid liquid. Then, on this account alone, the law of diminution of the sound proceeding directly ahead would be as follows. On advancing equal distances the intensity of the sound would lose equal percentages of the values it had on beginning each such distance. In other words, the intensity of the direct sound would diminish in geometrical progression as the distance increased in arithmetical progression. This same law has been found recently to hold for sounds under water in which many small obstructive patches occur. (For the recent development of submarine signalling see the article on "Sound Ranging.")

(vi.) *Diffraction.*—The spreading of sound behind obstacles, which is one form of diffraction, must also slightly weaken the main or direct beam of sound, beyond that due to other causes.

(vii.) *Fog, etc.*—We may now notice other states of the atmosphere which have been supposed prejudicial to the free passage of sound, but which Tyndall by an elaborate series of experiments found are not so. These include the presence of rain, hail, snow, and fog, which, he asserted, have "no sensible power to obstruct sound." Kelvin has shown that the temperature gradient, which forms the limiting condition of equilibrium of the air in a warm fog, is about half that in the limiting condition for fog-free air. This state is, of course, one in which the temperature falls as we ascend. Thus on theoretical grounds it should be expected that sound would pass with less loss in a fog than when the air is clear. This inference is fully borne out by Tyndall's prolonged researches off Dover.

It must be pointed out that the above statements, though correct, may be easily misunderstood and a false impression received as to what is really meant. Sound will usually travel better in a fog than in clear air, provided that the path in question is *wholly* in the *same* continuous bank of fog. That is to say, the sound must *originate in and be heard in the same continuous bank of fog.* If a sound originates in clear air and is to be listened for in clear air on the far side of an intervening bank of fog, then there might be large reflections at the first and second faces of the fog, and the sound might be much weakened or inaudible at the place where it was intended that it should be heard. Again, if the sound originates in one bank of fog and the listener is in a second bank of fog with clear air intervening, the sound might again be very faint or inaudible at the listener's position owing to the two reflecting surfaces interposed. Thus, though it is rightly asserted that sound travels well while *entirely in one continuous portion of fog,* the presence of a *number of banks of fog with intervening clear spaces,* all between the source and its desired destination, may prove very detrimental to the long-range transit of sound.

(viii.) *Zones of Silence.*—Many notices have appeared of late years calling attention to exceptional cases of sound propagation in which some explosion is heard in the vicinity, following which is a zone or annular space of *silence,* this being succeeded by a still more distant region in which the sound is again audible. Various theories have been put forward to account for this, such as (*a*) wind gradient, (*b*) temperature gradient, (*c*) more complex meteorological states of the atmosphere. The first alone may suffice to bring sound down again by total reflection at a height where a certain critical state obtains. The difficulty in settling what has been the cause of such phenomena on any particular occasion usually lies in the absence of qualified observers expecting an explosion and listening for it. Thus evidence collected afterwards may be biassed or in other ways untrustworthy.

III. Sound Producers

§ (24) General Survey.—We shall treat here the various producers of sound whether devised for music, signals, or other purposes, the first class, however, greatly preponderating.

It is often convenient to think of a musical instrument with its accessories (and certain parts of the performer himself) as divisible into three main portions. These are: (*a*) the *exciter,* or means of producing vibrations; (*b*) *the vibrating system*; and (*c*) the manipulative *mechanism* for the production of the scale, for expression, etc. Sometimes the vibrating system may be subdivided into a *vibrator of definite pitch* and a *resonator* which reinforces and otherwise modifies the sounds originated by the vibrator. In other cases there is a vibrator, not possessing a definite pitch of its own, but only favouring certain pitches, some one of which is made actual and definite by the resonator (in addition to its function of modifying the quality of the tone). The subdivision gives, in all, four portions of the instrument. Take as an example the case of the violin. In this the bow in the player's right hand is the exciter, the strings are the vibrators of definite pitch, the sound box and contained

air form the resonator, while the finger-board and the fingers of the player's left hand furnish the mechanism for the scale. The second form of subdivision is illustrated by the flute. Similar analyses may be supplied by the reader for the other instruments, special intricacies being referred to as we proceed.

Having thus briefly glanced at how the sounds may be produced, we now pass on to notice what are the characteristics which serve to distinguish one musical instrument from another. The chief of these are as follows:

(*a*) The range of pitches possible, or the *compass* of the instrument.

(*b*) The interval relation of the notes, or *scale* in use.

(*c*) The *power and delicacy* of the sounds producible.

(*d*) The *noises* accompanying the beginning or finishing of the sounds.

(*e*) The possible or inevitable change of intensity of the sounds while they last.

(*f*) The quality of the sound when established, by musicians termed *tone*.

(*g*) The possibility of sounding a number of notes together or the restriction to one at a time, *i.e.* the capacity for *harmony* or the restriction to *melody*.

Taking each of these seven points in order, we may illustrate them by the following contrasted pairs of instruments, of which one of any pair may be distinguished from the other on the above principles by hearing without seeing.

(*a*)	Harmonium	Concertina	Large, small compass.
(*b*)	Bugle	Cornet	Harmonic series, chromatic scale.
(*c*)	Trombone	Flute	Powerful, feeble.
(*d*)	Trumpet	Clarinet	Declamatory, smooth.
(*e*)	Harp	Violin	Sounds die away, sustained.
(*f*)	Oboe	French horn	Penetrating, muffled.
(*g*)	Violoncello	Piano	Melody, harmony.

The endeavour has been made to cite typical examples, but no doubt the discrimination between the members of each of the above pairs would depend partly upon other guiding factors, some perhaps too subtle to be recognised as such.

As to quality of *tone*, we may trace an almost continuous gradation from rough coarse bass quality, through dull, rich, full, or mellow tones, to those which are described as brilliant or even penetrating. Helmholtz has shown that the qualities of tone thus popularly described correspond to compound sounds in which the prime or fundamental sound of lowest frequency is accompanied by a retinue of other (and usually) fainter sounds of higher frequencies. All these sounds (including the lowest) are called *partials*. Musical sounds, as to their qualities of tone, may be scientifically divided into the following classes:

(i.) Those with the *full harmonic* series of partials; *i.e.* the sounds have frequencies as 1, 2, 3, 4, 5, 6, etc.

(ii.) Those with the *odd harmonic* series of partials; *i.e.* the sounds have frequencies as 1, 3, 5, 7, etc.

(iii.) Those with *inharmonic* partials; *i.e.* the sounds have frequencies *inexpressible by small whole numbers.*

(iv.) Those *without any upper partials.*

(v.) Those with harmonic partials, some of which near *fixed* pitches are *specially favoured* whatever the pitch of the prime.

Most musical instruments fall into the first of these classes; a few into the second; church bells, bars, and gongs into the third; tuning-forks into the fourth; while the human voice singing vowels stands alone in the last. (The question of the special tuning of keyboard and other instruments is treated in §§ 5-6.)

§ (25) Compasses of Chief Musical Instruments.—Table VII. (p. 699) gives the approximate compasses of the chief musical instruments classified both scientifically and musically. The notes given show the *true* pitches of the limiting sounds, and not necessarily those written for the instruments, which are often different according to the pitch of the instrument and the traditions of the profession.

§ (26) The Guitar.—This instrument has six strings, the three upper being of catgut, the three lower of silk covered with silver wire. The strings are plucked by the right hand, the thumb being used for the three lower strings, and the first, second, and third fingers for the three upper strings, the little finger resting on the body of the instrument. To obtain the different notes of the scale, the strings are pressed by the fingers of the left hand (or *stopped*) against the *frets* or little pieces of wood which cross the finger-board at the appropriate places. When the vibrating length of the string is thus reduced the frequency is increased and the note produced is accordingly higher. Thus, if the string's vibrating length is reduced to one-half the note is raised an octave, because its frequency is doubled. Again, if the length is reduced to two-thirds the note is raised a fifth, because the frequency is increased in the ratio of two to three. Similar remarks apply to the other notes of the scale. Thus every use of the frets raises the pitch above that of the open string. An account of the theoretical calculation of the frequencies of strings is given in § (52) (i.). Notes called *harmonics* may be obtained by touching the strings lightly with the fingers of the left hand instead of pressing them hard on to the frets. Thus, touching at the middle yields the octave, touching at either point of

trisection (*one-third* up or *two-thirds* up) yields the twelfth, touching at one-fourth up yields the double-octave. Hence (reverting to the typical subdivisions of a musical instrument) the thumb and three fingers of the right hand are the exciters, the six strings the vibrators, the sound box the resonator, while the fingers of the left hand and the frets are the manipulative mechanism for producing the scale. Since the notes are made by plucking they soon die away.

§ (27) THE MANDOLIN.—This instrument has four double strings, those of each pair being set to unison. The e'' strings are of catgut, the a' strings of steel, the d' strings of copper, and the g strings of catgut covered with silver wire. The notation for these strings is that shown at the top of Table VII. The strings are plucked with a plectrum (of tortoise-shell or horn), and *stopped* (that is pressed hard by the fingers of the left hand) upon the finger-board to produce the notes of the scale. H. Berlioz, the composer and writer on instrumentation, states that the quality of tone of the mandolin has a keen delicacy not possible on other instruments. This is probably due in part to the fine point of the plectrum used to initiate the vibrations. As in the case of the guitar the notes of the mandolin are soon damped out.

§ (28) THE HARP.—This is the important representative of stringed instruments played by plucking with the fingers. The double-action harp is tuned in C♭ and has seven pedals. Each pedal acts upon all the strings of a given name throughout the compass of the instrument. Further, each pedal may be used to raise the strings to which it applies by a semitone or by a whole tone at the option of the player. Thus, by the right use of the pedals before commencing a piece, the instrument may be arranged to provide any one of the fifteen major scales in the keys from C♭ to C♯, both inclusive. The melodic minor scales cannot be thus *set* by the pedals, since they require different notes in ascending and descending passages. The harmonic minor scales can be *set* in the twelve keys from A♭ to C♯ inclusive. The octave *harmonics* may be produced on the longer strings of the harp by touching them at the middle with the fleshy part of the palm of the hand and plucking with the thumb and two first fingers of the same hand. In the harp a separate string is provided for each note of the diatonic scale, and the pedals are the only accessories needed to give the accidentals or other sharps or flats.

The harp's quality of tone mingles well with the various brass instruments. The tone is also somewhat under the control of the player, since the strings may be plucked at various places and in a variety of ways.

It is to be noted that the lower notes of the harp, being made by long and heavy strings, are fairly powerful, and not so quickly quenched as those of the guitar and mandolin. Indeed, they often need to be *damped* or *checked by the hands*. Still, since they are plucked, they must die away in time, having no supply of energy to sustain them. Further, the fact that the string is excited by plucking leaves the string free to follow the vibrations natural to it, and this gives a certain characteristic quality to the tone of such instruments. This is due to the fact that the higher partials of a thick string are slightly *inharmonic* and will *so sound* in the case of a plucked or struck string, whereas in a bowed string all the vibrations are forced by the bow into the strictly harmonic frequencies.

§ (29) THE PIANOFORTE.—In this instrument [1] the strings are of cast-steel wire, at most parts of the compass being in duplicate or triplicate, but at the bass end single and over-spun or wrapped with copper or other wire. In the grand pianos the strings are arranged horizontally to admit of greater length and correspondingly increased power. The sounds are excited by the felt-faced hammers *striking* the strings, then rebounding and leaving them free to vibrate. The hammers are set in motion by the *action* (or train of mechanism) intervening between them and the keys which form the familiar keyboard. The action also lifts a felt-covered *damper* from the string when the key is struck, holds it off while the key is held, and replaces it to terminate the sound when the key is released. The depression of the right pedal takes the dampers off all the strings and thus allows a considerable increase of loudness and change of quality of tone through the sympathetic resonances of strings which are not struck. The use of the left or soft pedal either (*a*) restricts the vibrations to a single wire for each note, (*b*) muffles the sound by the interposition of a layer of felt, cloth, or other soft material between the strings and the hammers, or (*c*) obtains a softened effect by shortening the stroke of the hammers. Along the upper part of the frame (in the upright piano) there is the wooden wrest-plank carrying the tuning pins into which the strings are fastened, their lower ends being secured to the hitch pins placed along the lower parts of the frame. The strings rest upon hard-wood bridges glued to the sounding-board. This is often of spruce fir. It is slightly convex towards the strings, and is strengthened by bars of pine glued to its inner concave side.

Having thus outlined the construction and action of the piano, let us turn to the factors upon which depend its retinue of partials and, consequently, the quality of its tone.

[1] See also "Pianoforte, The."

TABLE VII.—COMPASSES OF CHIEF MUSICAL INSTRUMENTS

	INSTRUMENTS SCIENTIFICALLY ARRANGED.		Compass (C, C c c′ c″ c‴ c^iv; 8ve bassa, loco, 8ve)	MUSICAL DESIGNATIONS.
STRINGS.	Plucked	Guitar	E — A	Unusual.
		Mandolin	G — E	
		Harp	C♭ — F♯	
	Struck	Pianoforte	A — A	For Solos and Accompaniments.
	Bowed *	Violin	G — G♯	The "Strings" of the Orchestra.
		Viola	C — A	
		Violoncello	C — A	
		Double Bass	G — G	
WIND.	Metal Reeds	Harmonium, American Organ	C — C	Keyboard Instruments with separate Reed or Pipe for each Note.
		Concertina	G — C	
	Do. with Pipes	Reed Stops of Organ	C — C	
	Pipes without Reeds	Flute Stops of Organ	C — C	
		Flute (in D)	C — C	The "Wood Wind" of the Orchestra and of Military Bands.
		Piccolo (in D)	C♯ — A	
	Double Cane Reeds with Conical Pipes	Oboe	B — F	
		English Horn	E — B♭	
		Bassoon	B♭ — B♭	
		Double Bassoon	D — F	
	Single Cane Reeds with Cylindrical Pipes	Clarinet (B♭)	D — F	
		Alto Clarinet (E♭)	G — B♭	
		Bass Clarinet (B♭)	D — F	
	Hyperbolical Pipes with Cupped Mouthpieces; Human Lips as Reeds	French Horn (F)	F — F	The "Brass" of the Orchestra (and used also in Military Bands).
		Bach Trumpet (E♭)	A — E♭	
		Trumpet (F)	B — B♭	
		Alto Trombone (E♭)	A — G	
		Tenor Trombone (B♭)	E — D	
		Bass Trombone (G)	C♯ — G	
		Bass Tuba (C)	C — E	
		Soprano (Cornet) (E♭)	A — E♭	Ordinary Valved Instruments of "Brass" Band and of Military Band.
		Cornet (B♭)	E — B♭	
		Tenor Horn (E♭)	A — E♭	
		Euphonion (B♭)	B♭ — B♭	
		Bombardon (E♭)	A — E♭	
		Double Bass (B♭)	E — B♭	
	Human Voices	Soprano	D — A	"Soli" and Chorus.
		Contralto	G — D	
		Tenor	C — G	
		Bass	F — C	
	Membrane	Kettledrum	F - - - - - F	Percussion Instrument of Definite Pitch.

* When played *pizzicato*, these fall under the heading of plucked strings.

As we have seen, the exciters are the hammers, the vibrators of definite pitch are the strings (one or more for each note), and the sound-board is the resonator. Thus the blow of the hammer, at a spot chosen by the maker, imparts a transverse velocity to a small portion of the string. The state of things produced there runs to and fro along the string. This leads to a particular type of vibration according to the *form and softness* of the hammer face and the *place* at which it strikes the string. A sharp hard edge to the hammer would favour very many partials and develop a tinkling or piercing tone; the rounded soft face of the hammer in actual use gives a justly prized mellow tone. Again, if the string were struck at a seventh of its length from one end, it is impossible for this point to be a place of no motion. Consequently the seventh partial is not encouraged, for it would correspond to the vibration of the string in seven sections with one of its *nodes* (or places of no motion) at a seventh from one end. But this remark as to no motion at the node applies only to the vibration for which it is a node. Hence, though the hammer by striking at a seventh would not encourage the seventh partial, this partial might still creep in. In some pianos the striking place is between the seventh and the ninth of the length of the strings from one end, varying in different parts of the compass. The time of contact of the hammer with the string may be of the order of half the period of the prime or fundamental of that string. Thus for the middle C (of, say, 264 per second) the time of contact might be less than one five-hundredth of a second, perhaps as low as one six-hundredth of a second.

We have seen that the character of the vibrations of the piano string itself depends largely upon the nature of the hammer and the position of its blow. It must be noted, further, that the character of the vibrations depends also on the thickness and material of the strings. Thus, very thick strings of rigid material would be too stiff to form very high partials, since they correspond to the vibration of the strings in many small segments and involve considerable bending. And for the same reasons any moderate partials would be inharmonic and would so sound.

But when the type of the vibrations of the string itself is finally settled, is the instrument committed to a definite effect as to quality of tone? Certainly not. Another very important factor remains to be examined. This is the sound-board and its bridge upon which the string presses.

The importance of the sound-board may be seen thus. If a string isolated from the sound-board vibrated precisely as a pianoforte string does, very little sound would be heard. The reason is that the air would flow and reflow round the string and be scarcely disturbed far away. A sound-board, on the other hand, cannot vibrate without setting in motion a large body of air, thus starting waves of sound that are audible all over a large room. Accordingly we must ask, not merely what vibrations does the string itself execute, but also what vibrations does the string force the sound-board to execute? Of course, the latter vibrations depend upon the former, but may differ from them in essential respects.

If the vibrations of the string itself are precisely all that could be desired, then it is the duty of the bridge and sound-board to reinforce them *without change* of character and convey them to the air. If, on the contrary, the vibrations of the string are defective in any way, then on the bridge and sound-board is imposed the double duty of reinforcing *and improving* these vibrations so that the desired quality of tone is received from the instrument as a whole.

The exact types of complex vibration which are most acceptable as forming the ideal musical tone are to some extent matters of opinion or controversy. Still more difficult is it to assign the precise dimensions of the mechanical details of hammer, string, bridge, and sound-board which would secure their production. Probably the nature of the wood of the sound-board and its seasoning are very important. Possibly also the exact dimensions of a sound-board must be varied according to the particular nature of the specimen of wood used for it. The various points here briefly referred to have been under investigation for some time, but no final solution of such intricate problems can be expected at any early date. In matters of this kind we no doubt owe much to the refined taste and instinct of those concerned in the making of pianos, and to their accumulated experience during two and a half centuries.[1]

§ (30) THE VIOLIN FAMILY.—The instruments of this family now in use in England are four, namely, the violin, the viola, the violoncello, and the double-bass (see *Figs.* 4A, 4B, and 4C). The violin has its strings of catgut tuned to e'', a', d', and g, the latter being covered with silver wire. The notation here used is that at the top of Table VII. § (25). The viola has its strings tuned to a', d', g, and c, all being thus a fifth lower than those of the violin. The violoncello is pitched an octave below the viola, its strings being accordingly tuned to a, d, G, and C. The double-bass may

[1] For research work on piano vibrations see papers by G. H. Berry, *Phil. Mag.*, April and October 1910 and July 1911; also one by C. V. Raman, *Roy. Soc. Proc.*, April 1920, xcvii. 99-110.

have three strings or four, and they may be tuned to fourths or fifths, the earlier alternative in each case being that in more general use in England. Thus the three-stringed bass may be tuned in fourths to G, D, and A_1, or in fifths to A, D, and G_1.

Except for the strings, their tuning, and the corresponding sizes of the instruments, they have much in common. Thus many of the following remarks apply to all of them. The bow with its rosined hairs is the exciter, the strings are the vibrators of definite pitch, the sound-box is the resonator, while the finger-board, *free from frets*, allows the scale to be played by the fingers of the left hand. The vibrating portion of the string reaches from the bridge to what is called the *nut*, the little cross-piece of ebony or wood between the neck and head or peg-box. When the left hand is farthest from the bridge it is said to be in the first position. Then, the first, second, and third fingers when used in turn to *stop* the string (*i.e.* to press the string against the finger-board) give the notes of the diatonic scale required to bridge the interval between one open string and the next. Thus from the fourth or *g*-string of the violin, these fingers give in turn the notes *a*, *b*, and *c′*. The next open string is *d′*, on which the fingers stop the notes *e′*, *f′*, and *g′*. The fourth finger may be used on any of the three lower strings to give a note in unison with the open string above. When the hand is shifted nearer the bridge so that the first finger occupies the place formerly held by the second finger, the hand is said to be in the second position. The third, fourth, fifth, sixth, etc. positions are also used, the odd being easier than the even positions. Each position has its advantages as regards the notes commanded on a given string, either for the sake of (*a*) smoothness of phrasing in certain passages, or (*b*) the special quality of tone of the string on which such notes are now played. For each string, differing in thickness and mass from its fellows, has its own tone and other individual characteristics.

FIG. 4A. Violin.

FIG. 4B. Violoncello.

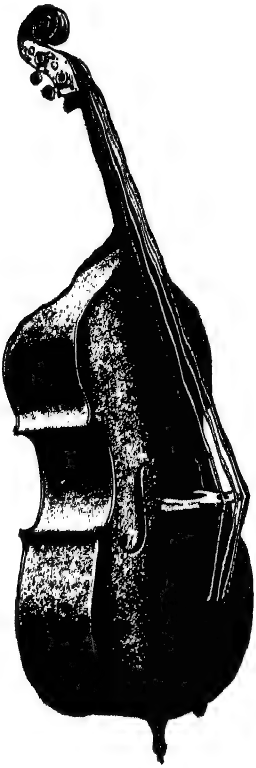


FIG. 4C. Double-Bass.

The fingering of the violin in the first and third positions, in illustration of the above remarks, is shown in *Fig.* 5. In the third position the notes indicated by $\overset{4}{0}$ are played by the fourth finger stretched out a note higher and only *touching* the string lightly at its "centre," thus eliciting the *octave harmonic*. When the term centre is used here it must be interpreted, not necessarily as the point equidistant from bridge and nut, but as that point which divides the string into two segments which vibrate in the same time. This is often not quite the same as the former, because the strings are usually slightly tapering from end to end.

Let us now examine more closely the action of the violin and the rôle played by its various parts. We may thus learn something of the qualities a fine instrument should possess, and also of what its technique demands. The stroke of the bow pulls the bowed part of the string aside at *its own speed*;[1] this part then springs back, usually at a different speed. Further, when the bow properly rosined is applied with due pressure at the right place

[1] Surmised by Helmholtz and recently (1914) established by C. V. Raman. (See *Science Abstracts*, A, 1915, xviii. 87.)

on the string and moved at the appropriate speed at right angles to the length of the string, then the full series of partial tones is elicited, and with a special grading of their intensities as follows:

Relative frequencies of partials	1,	2,	3,	4,	5,	6,	etc.
Intensities in violin tone	1,	1/4,	1/9,	1/16,	1/25,	1/36,	etc.

Or, in words, the intensities of the partials of a well-bowed string are *inversely as the squares of their frequencies.* This highly specialised composition constitutes what may be called violin tone. But these partials do not make any immediate appeal to the ear. The vibrations of the string first affect the bridge (of maple); its feet press upon the pine belly and set that in sympathetic vibration. The belly sets the ribs (of maple) and back (maple or sycamore) in vibration, and thus also the air contained in the sound-box, which passes in and out at the two *f*-holes. The belly is strengthened by the bass bar, which is inside it and lies beneath the *g*-string. Nearly under one foot of the bridge (on the e''-string side), the belly is supported by the sound-post, a little round piece of pine which is held in place by the pressure between back and belly.

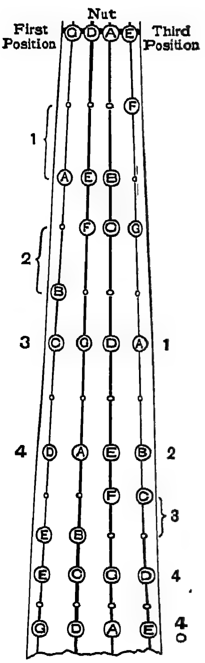


FIG. 5.
Fingering of Violin.

It seems highly probable, therefore, that the main portion of the sound heard is that proceeding from the vibrations of the belly, the back, and the air between them. Further, since these structures are of complicated shape with asymmetrical strengthening pieces, it is not likely that they can follow precisely the vibrations conveyed to them from the string by the bridge. Indeed it has been shown experimentally that the vibrations are modified in character as they pass through the series of moving parts: string, bridge, belly, and air between it and the back. And this is only what may be expected from the practical consideration that the value of the instrument lies in the *sound-box* that *modifies* the vibrations in a certain individual manner, and *not* in the *strings* that *originate* them. The fine old violins have also a special varnish which preserves and perhaps enhances their tone.[1]

Let us now consider what the violin requires from the performer. The fingers of the left hand have to stop the strings so as to produce the notes in the *intonation desired* according to what other instruments are used in concert with the violin. Further, this task must be performed without any help from frets, which are not present on the keyboard. But this leaves the possibility of gliding the finger from one stop to another, thus introducing the grace called *portamento,* which must be used with discretion. The finger holding a stop may also be rocked or rolled, thus producing an effect called the *tremolo.* It may be executed rapidly to intensify dramatic expression, or slowly to impart tenderness to a pathetic melody. Thus, except for the portamento and the tremolo, only *accuracy* (in stop and time) is demanded from the left hand. But from the right hand are required not accuracy only but the production of all the tone and expression of which the violin is capable. The bow should always move perpendicularly to the strings; the place of bowing must be chosen according to the effect desired, near the bridge for loud passages, farther away for softer ones. Again, with each such place must be associated the corresponding pressure and speed. The strings of the harp are plucked, those of the piano struck, and the vibrations of each die away according to the inexorable law of such instrument, over which the player has no control. The vibrations of a bowed string, on the contrary, may be sustained of constant intensity, or may be gradually increased or decreased in intensity at the will of the performer. Indeed, in this respect, everything is possible on the violin, though correspondingly difficult.[2]

§ (31) METAL REEDS WITHOUT PIPES.—The harmonium, the American organ, and the concertina present examples of this class of vibrator, the reeds in question being called *free reeds.* A reed or tongue of this type is a thin oblong metal plate or strip fastened at one end to a block in which there is a hole behind the tongue of the same shape. In some examples of these reeds the tongue, when at rest, closes the hole in the block except for a very fine clearance at its margin. When in motion the tongue alternately opens and (nearly) closes the hole in the block. Now, it is a result of theory that, the more sudden the discontinuity of any periodic motion,

[1] For research work on the violin see *Phil. Mag.*, Aug. 1909, Sept. 1910, June 1912; also *Bulletin*, No. 15, Indian Association for the Cultivation of Science, 1918.

[2] For a research on these points see *Phil. Mag.*, May 1920, xxxix. 535-536.

the greater the relative importance of the high partials which are present in the corresponding sound. Thus, the more sudden the action of the reed, the more obtrusive are the high partials in the sound, and the more cutting or grating is the quality of tone. Further, since in the instruments now under notice there is no pipe whose resonance might modify this quality, it remains of that cutting character which soon palls.

To sharpen the pitch of a reed, a little is scraped off the *tip*. This, while not changing the spring, diminishes the mass at the end where its effect is greatest and so increases the frequency. To flatten the pitch, a little is scraped off the *root* of the reed. This, while not changing the effective mass, weakens the spring and so decreases the frequency.

In the harmonium the bellows *forces* the wind through the reeds, while in the American organ the wind is *drawn* through the reeds into the bellows. The English concertina is tuned to mean-tone and has separate notes for D♯ and E♭, also for G♯ and A♭.

§ (32) REED PIPES IN THE ORGAN.—These differ in two respects from the reeds of the harmonium, *first* in having a pipe to modify the quality of tone produced by the reed, *second* in the manner of their tuning. It is also to be noticed that the reeds are of two kinds, free reeds like those in the harmonium, and *beating* reeds which are too large to pass into the opening with which they are associated. They therefore bend on to the opening like a covering flap until they have nearly closed it. Such a reed is adjusted, or *voiced*, so as to close with a rolling motion, and thus gradually cover the aperture. The harshness of quality consequent upon a sudden discontinuity is thus obviated. The free reed of an organ pipe is tuned by a wire clip which grasps it near its root. The shorter the vibrating part of the reed the higher the pitch. The beating reed has a wire pressing upon it near the root, and by its adjustment the reed is tuned. By varying the design of the reeds and the shape of the pipes with which they are used, various typical tone qualities can be obtained or approached, thus imitating or suggesting the various orchestral instruments.

In these *reed stops* (or sets of pipes of a given type for several octaves) we may note the following parts and functions. The blast of wind is the exciter, the reed of definite pitch is the vibrator, and the pipe with which it is associated is the resonator modifying the quality of tone produced.

§ (33) ORGAN PIPES WITHOUT REEDS.—These constitute the flute, or flue, stops of the organ. They are usually parallel and always have a typical opening called the *mouth*. Above the mouth is the *speaking length* of the pipe, below it is the *rack length*, or portion of smaller tube by which the pipe is put in connection with the wind chest from which it is blown. Metal pipes are usually of circular cross-section, wood pipes are usually rectangular in cross-section. The flute pipes also fall into two main classes according as they are *open* or *stopped* at their top end.

Let us inquire how these pipes come to utter their musical sound. (See *Fig.* 6.) Suppose the pipe is in position, the air in the wind chest at the right pressure, and the key is depressed. Then the blast of air issues from its slit-like opening at the lower part of the mouth of the pipe and passes up in the

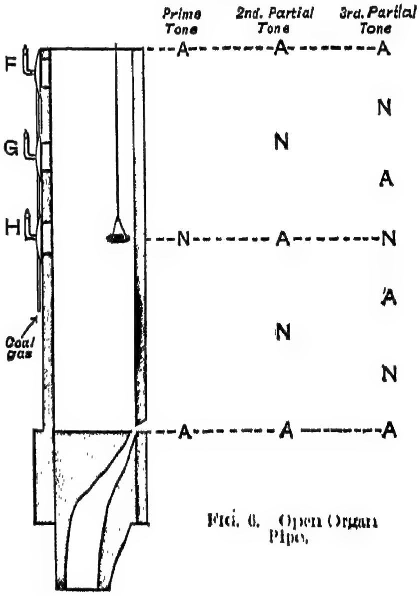


FIG. 6. Open Organ Pipe.

form of a thin sheet or blade and strikes the *lip* or upper edge of the pipe's mouth. It may pass chiefly outside the pipe and so make a rarefaction inside or it may pass chiefly inside and so make there a compression. Whichever state is first produced just inside the mouth advances to the top of the pipe, is there reflected, and returns to the mouth. On reaching the mouth this reflected disturbance reacts upon the unstable blade-shaped stream of air still passing upward at the mouth. Accordingly this stream is easily pushed or drawn aside, and by this very deflection encourages that state which already exists. Thus the wind maintains the sound of the pipe. But its pitch or frequency is fixed by its speaking length, the nature of the reflection at the top, and the speed of sound in the air of the pipe.

(i.) *Open Pipes.*—If the pipe is open then a compression arriving there will be reflected as a rarefaction, for the two opposite states are required to keep the pressure practically constant, as must be the case at an open end. Hence, if a compression starts up from the mouth, it is changed into a rarefaction after its ascent, and after its descent it is again reversed in state by reflection at the mouth, so starts up again as a compression. That is, the original state of things is repeated after *twice* traversing the speaking-length of the pipe. Consider now the case of a stopped pipe. A compression arriving at the top end is reflected as a compression (the double compression being easily supported by the stopped end), descends to the mouth, is there reflected as a rarefaction, and passes up and down the pipe again in that state. Then, by reflection at the mouth, it is changed to its original state of compression. And this occurs after traversing *four times* the speaking-length of the pipe. Thus a stopped pipe has about double the period and half the frequency of an open pipe of the same length. In other words, stopping the pipe lowers its pitch about an octave. All the above remarks apply to the prime or fundamental of the pipe. Let us now inquire what are the pitches of the other partials possible to the pipe. Take first an open pipe as shown in *Fig.* 6. To prepare for the upper partials we may with advantage go into a little closer detail in the prime itself. Thus the mouth and the upper open end are both *antinodes* or places of freest motion. These are accordingly marked A in the first column of *Fig.* 6 under the heading prime tone. But, half a period after a compression leaves the mouth to ascend, a rarefaction leaves it also and starts up the pipe. Now, at this very instant a rarefaction, derived from the previous compression, is starting down from the top. Thus, these two rarefactions will meet at the middle of the pipe, and while (by their opposite direction) annulling all motion they will produce a double rarefaction. This middle point is therefore for the prime a *node*, or place of no motion but greatest change of pressure. It is accordingly marked N on the *figure*. If the pipe were overblown it could sound its second partial tone. The corresponding state is as shown in the column under that heading. Thus, there are again antinodes at the open end and mouth, but now an antinode in the middle also, nodes occurring midway between the antinodes as before. The wave-length of the vibration is accordingly half that for the prime and the frequency, in consequence, is doubled. The next column shows the state of things for the third partial, in which the wave-length is a third that for the prime and the frequency is trebled. In *Fig.* 6 a little tambourine with sand is shown at the middle, by which it may be demonstrated that there is no motion at the nodes and much at the antinodes. The gas jets F, G, H show that there is a fluttering of pressure at the nodes.

Now, when the pipe is supplied with wind of the appropriate pressure it sounds the prime most prominently, but also the other partials too, their intensity falling off as we ascend in the series. A set of pipes intended for use at any given pressure must therefore be so *voiced* (or adjusted at their mouths) as to respond readily to blowing, and start their notes as promptly as possible. If the stream of air were so directed as to fall outside the pipe always, or inside it always; then in each case the pipe would not speak. It must be so directed and of such unstable nature as to be easily pushed out or drawn in by the first feeble pulse that has travelled up and down the pipe. Further, this stream of air must be of such nature as to encourage, not the prime only, but each of the other desired partials also, and each in the degree desired. But the pipe has a great effect upon the final result whatever the stream of air is favouring. Thus, with a narrow open pipe, all the partial tones are well in tune, that is the frequencies possible to the pipe are almost exactly in the ratio 1, 2, 3, 4, 5, 6, etc. Now when the blast is on, all the partial tones actually sounded must be precisely in tune as just mentioned. Hence, on blowing such a pipe these partial tones readily respond and the tone is fairly bright because many partial tones are present. Again, consider an open pipe, but this time a wider one (and perhaps of wood), the possible partials have frequencies nearly as 1, 2, 3, 4, 5, 6, getting more out of tune as we ascend the series. Accordingly it requires more effort to induce these higher partials to speak in tune, which they must do under the influence of the blast if they are to speak at all. The result is that these higher partials speak more feebly in a wide pipe than in a narrow one, and the tone is consequently mellower or less bright.

(ii.) *Stopped Pipes.*—Turning now to a stopped pipe, we must always have a node at the stopped end and an antinode at the mouth. For the prime tone there is no other node or antinode in the pipe, which accordingly has a length about a quarter of the wave-length speaking. For the next partial we should have the state of things represented by ANAN; beginning at the mouth and denoting by A an antinode and by N a node. The length of the pipe is accordingly three-quarters of the wave-length now in use. In other words, the wave-length is now one-third that for the prime and the frequency is trebled. Similarly for the next higher partial we should have the scheme ANANAN. This shows that

the pipe is five-quarters of the wave-length now in use, whose frequency is accordingly five times that of the prime. Thus, for a stopped pipe the frequencies of the partials possible are as 1, 3, 5, 7, etc., *i.e.* as the odd numbers only. With the stopped pipes again (as with the open ones), the narrowness of the pipes favours the upper partials and produces a brighter tone, while widening discourages them and gives a duller tone. But, whether wide or narrow, the stopped pipes have a tone quality distinct from the open ones because of the odd partials only being present instead of all. The tone of stopped pipes may be described as somewhat hollow, the wider stopped pipes have a dull quality of tone, and when very low are said to be soft and powerless.

The tuning of open flue pipes when of metal and cylindrical is accompanied by shifting a sleeve which slides on the upper end. Pushing the sleeve up lengthens the pipe and flattens the pitch, pushing it down shortens the pipe and sharpens the pitch.

Wooden open pipes are tuned by bending a metal piece which shades the top. The pitch is flattened by lowering this flap, because the open end is then more effectively shaded and this is equivalent to lengthening the pipe. For the *virtual* open end, as regards the waves of sound, is always beyond the actual end, and the less open the end is the greater the discrepancy between the two. To sharpen the pitch the flap or shade is raised and the pipe virtually shortened.

§ (34) The Flute and Piccolo.—In these instruments (and fifes) we have a pipe parallel, or nearly so, open at one end and pierced with a special mouth-hole near the other end. They are thus comparable to the flute pipes of the organ as to the manner of maintaining their sound, though the lips and chest of the flautist replace some corresponding mechanism in the organ. The air blast may be regarded as the exciter, while the vibrating system comprises (*a*) the blade-shaped stream of air which passes from the player's lips and strikes the sharp edge of the mouth-hole, and (*b*) the column of air within the cylindrical pipe. The pitch is decided and kept steady by the length of the air column whose vibration causes the stream of air alternately to enter or pass over the mouth-hole. The manipulative mechanism for the production of the scale consists of the holes and keys along the side of the tube which act by regulating the effective length of the tube in use. It is this length which governs the time of passage of a pulse to and fro in the tube. In a very simple form of flute or piccolo there are only six finger holes and six keys, the former being open except when closed by the fingers, the latter closing holes except when opened by the fingers. In such a flute the uncovering of the holes in order beginning at the right-hand or open end would usually give the diatonic scale of D. If, however, the tube is lengthened and other keys provided, the chromatic scale may be continued down to C natural. By some musicians the flute is then said to be in C, although the holes still correspond to the scale of D. In Table VII., p. 699, flute in D

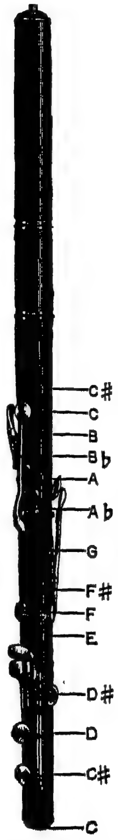


Fig. 7.—Concert Flute, Pratten's Cone.

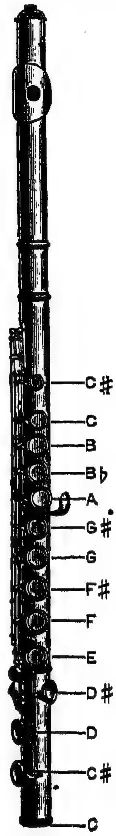


Fig. 8.—Silver Concert Flute, Boehm Cylinder.

is inserted although of this type. In the Boehm flute the holes are all covered, but the positions of these key-holes correspond to all the finger-holes and key-holes in the simpler arrangement. The two flutes are shown in *Figs.* 7 and 8. The musical notes indicated against the open end and the other holes give the pitches sounded in the lowest octave when each place is in turn the first opening counting from the mouth-hole. In other words, the place in question is open but everything else is closed between there and the mouth-hole. The higher octaves are obtained by over-blowing and slightly different fingering

The flute has the full series of partials since its tube is parallel and open at each end, but the partials do not usually extend far. Thus, the quality of the flute's tone is very mellow and pleasing, being somewhat dull in the lower register if sounded softly, but brighter in the middle part of the compass. The partial tones of the flute are in some cases almost limited to the prime and octave. In attempting to obtain low notes *loud*, the octave (or second partial) may be much louder than the prime, and there is difficulty in keeping the prime going at all.

The piccolo is about an octave above the flute, but in other respects similar. The lower part of its register is poor in quality and is not much used, the flute giving better notes in this part of the compass. Also the higher notes of the piccolo, if loud, are somewhat harsh and need toning down by a sufficient accompaniment. But the piccolo, judiciously used, can give an incisive brightness not otherwise obtainable. It can also continue a melody beyond the range of all other orchestral instruments.

The flute and piccolo are the most agile of the wood-wind instruments, being capable of rapid staccato passages which would be difficult and ineffective on reed instruments. This is in great measure due to the possibility of double and triple tonguing on the flute, which is impossible with a reed. To grasp this point we must note that the ordinary method of producing or *tonguing* detached notes on the flute is by articulating the syllable *too*, or something which approximates closely to it. Now this is done by sharply withdrawing the tongue from between the lips and emitting a puff of air. But if this syllable *too* were required for every detached note, however rapid the music, only a confused scuffle would follow. For the tongue would have to perform two motions, backward and forward, for every note uttered. This difficulty is obviated by pronouncing different syllables alternately, say *too* with the backward stroke of the tongue, and *koo* with its forward stroke, a musical *note* being obtained at each stroke. Various syllables are used according to individual choice, clearness and equality of articulation being the aim. It is easily seen that double-tonguing offers the same advantage to the flautist that up-and-down bowing does to the violinist, a note being obtained at each stroke, instead of only one note for both strokes, as in single tonguing on the flute, or using all down-strokes with the violin bow.

§ (35) THE OBOE AND BASSOON.—These are the soprano and bass instruments of the same family. Each is played with a small double-cane reed, and each has a conical tube. It can be shown mathematically that such an instrument (a cone closed by the reed at its vertex) must have the full series of partials. And it is found by experiment to be the case for the oboe. Indeed, some of its high partials are very powerful, and therefore make the tone quality thin but penetrating. The oboe takes so little air that the player is practically holding his breath while sounding a continuous

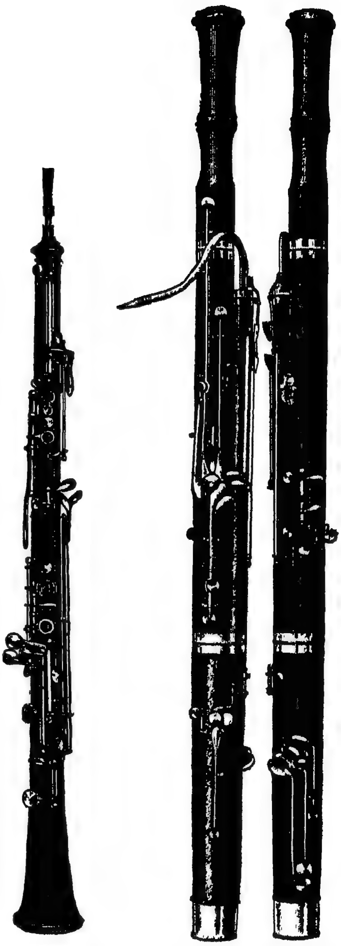


FIG. 9.—The Oboe. FIG. 10.—The Bassoon.

note, and needs rest not to take breath but to expel that which could not be passed through the reed. With these instruments the air blast is the exciter, the reed and its associated tube form the vibrator, and the somewhat complicated system of holes and keys supply the mechanism for the production of the scale. This is accomplished by changing the effective length of the tube from the reed to the first opening.

The conical tube of the oboe is straight,

while the bassoon, about two octaves lower, has a bent tube. The bell of the bassoon is at the upper end, the other or reed end being near the bell, but lower, while the conical tubing passes down from the reed to the concealed bend at the lowest part of the instrument and thence straight up to the bell. Both instruments are shown in *Figs.* 9 and 10.

The English horn or *Cor Anglais* may be regarded as an alto oboe, but the bell is rather different in shape, and the tone is quite characteristic and, in certain cases, extremely effective (see *Fig.* 11).

FIG. 11. Cor Anglais.

All the instruments of this family (like the flutes) yield the octave by overblowing. By this device, therefore, combined with suitable fingering, the higher octaves are obtained.

§ (36) THE CLARINET.—This instrument is very special from the scientific standpoint and justly prized by musicians for its fine tone. It is played with a single beating-reed of cane; the tubing is parallel, except for a small part near the mouthpiece and the end which forms a bell. Thus, since the reed must count as a closed end, the clarinet has the odd-numbered partials most prominent, although the even ones are not entirely absent. It is this, together with the special behaviour of the reed in the mouth of a skilled performer, that gives the tone of the instrument its quality so characteristic of the clarinet. Further, it must be noted that the instrument (because of its *odd* partials) overblows the *twelfth* (instead of the octave, as in the flute above). In other words, since the strong partials have frequencies, as 1, 3, 5, 7, etc., that of triple frequency is the next note to the prime, and so is got by overblowing. And this note is the twelfth of the prime in the musical method of reckoning by counting the notes of the diatonic scale. Accordingly, on the clarinet a sufficient number of holes and keys must be provided to give the chromatic scale over the interval of a twelfth from low E up to B. This requires at least 19 holes and keys, but it has been found preferable to provide 7 holes and 13 keys. These are, as shown in *Fig.* 12, the letters against each hole or key indicating the notes sounded when the place in question is open and all else is closed between there and the reed.

Since the fingering is somewhat complicated in the clarinet it is found preferable to keep it within the limit of keys without more than two flats or sharps in their signatures, at any rate for rapid passages. This necessitates the provision of clarinets in various keys, those in common use in the orchestra being in C, in B♭, and in A. Thus, if a piece of music required performing in the key of E♭ (three flats in the signature), by directing the use of the B♭ clarinet the music would be written in F (one flat only in the signature). For the fingering that produces F on the C clarinet will produce E♭ (a tone lower) on the B♭ clarinet, since the instrument itself is a tone lower by reason of its extra length of tubing. Again, if music in the key of E (four sharps in the signature) were to be played on the clarinet, the music would be written in G (one sharp in the signature) and assigned to a clarinet in A. Then, since this instrument gives A for the note written and fingered as C, it will give E for that written and fingered as G. Of course the C clarinet could be used for the five keys B♭, F, C, G, and D; but the B♭ clarinet would be better for the first of these, about equally good for the second, while the A clarinet would be better for the last. But apart from the case of fingering which has just been alluded to there is the distinctive quality of tone of each instrument to be considered. Thus the C clarinet is in the upper parts somewhat unsympathetic, while the A clarinet is duller than the B♭ instrument, which is considered the richest and fullest in tone.

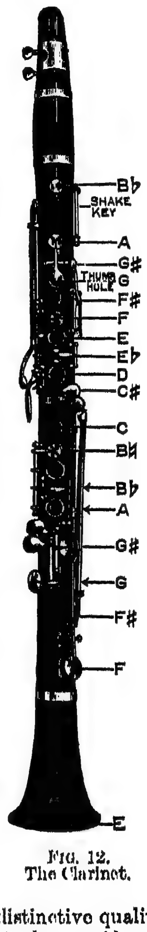


FIG. 12. The Clarinet.

It may be noted that the clarinet is specially capable of the most delicate gradations of *piano* and *pianissimo*. Further, in common with all the other reed instruments, the clarinet is heard at its best in legato phrases. Staccato playing on the reeds is ineffective

in rapid passages, as no double-tonguing is possible (as in the case with the flute).

§ (37) THE FRENCH HORN WITHOUT VALVES.—We now commence the notice of a class of instrument very important musically and of high scientific interest, namely, the *brass* of the orchestra and other bands. All these have hyperbolical pipes and cupped mouthpieces, the human lips forming the very special double-reed. To produce any desired note, that is possible with the length of tube in use, the lips (at the right tension) are applied to the mouthpiece, the tongue is suddenly withdrawn from them and a sound like *too* articulated. This sound is steadied in pitch and much increased in power by the tube. Thus the blast of air is the exciter, the human lips and the tube together forming the vibrator. In the above respects all the brass instruments are alike. They differ, however, very widely in the mechanism for the scale, in compass, in power, and in many other details.

All French horns have very long tubing coiled in a circular form and ending in a large bell, and are played with a long narrow tapering mouthpiece which favours the production of high notes of a soft quality of tone (see *Fig.* 13). In playing, the right hand is placed

FIG. 13.—French Horn without Valves.

inside the bell to give the tone a slightly muffled or veiled quality. The hand can also be changed somewhat in position so as to alter the pitch by various fractions of a tone, quarter-tone, semitone, three-quarters tone, etc. This changed position of the hand involves also a change in the degree of muffling.

In common with all the other brass instruments French horns have the full series of partials. This is due to the fact that a cone closed at the vertex gives this full set, and that all the brass instruments are quasi-conical. The curvature is to correct for the bell and mouthpiece, etc. By a suitable pinch of the lips those different partials may be elicited in turn as the *open notes* of the instrument, those which are lower than the note required being suppressed. In this way the various possible frequencies from 1 to 16 may be sounded. These notes for a horn in C may be represented in ascending order as follows:

[C]	c	g	c'	e'	g'	($b'\flat$)	c''	d''	e''	(f'')	g''	(a'')	($b''\flat$)	b''	c'''.
[1]	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16.

The notation here used is that at the top of Table VII., p. 699, giving the compass of the various instruments. The first note or *pedal* is shown in square brackets and is very seldom used musically. Indeed, it is not strictly in tune, but is somewhat displaced in pitch in order to bring the other notes that are used into more exact tuning. Again, the 7th, 11th, 13th, and 14th notes are in round brackets to indicate that they are not in exact tune with the notes shown by $b\flat$, f, and a, nor indeed with any notes in the tempered scale in use. But, if the instrument is a good one, the notes in question have the frequencies shown by the numbers 7, 11, 13, and 14. It is seen that the series of open notes give some approximation to a diatonic scale in the top octave but not in the lower ones. In the original form of the horn (which is very ancient) no valves were provided, and the fist had to be used to modify the open notes if an approach to the ordinary scale were required in parts of the compass where the open notes left large gaps. For this reason the French horn without valves is often called a *hand*-horn. But when the instrument was used without valves there were still gaps in the lower parts of its compass where the open notes are a musical fourth or fifth apart. To meet this deficiency horns were used in various keys by the addition of coiled tubes, called *crooks*, near the mouthpiece. The crooks were of different lengths, so as to put the horn in A, A♭, G, F, E, E♭, etc. Thus, by having some horns in one key and some in another, a composer could get all the notes he required. Such crooks are still used on the French horns with valves, because, by putting the horn in a suitable key, more open notes are obtained. In order to tune the horn some parallel tubing is provided, of which one piece like a U can be drawn out at pleasure. This is called a tuning slide. When no valves were used the horn would have a characteristic of its own, owing to its scale limitation, apart altogether from its special quality of tone. The same thing is noticeable to-day in bugle calls and in fanfares on trumpets without the use of valves.

§ (38) French Horn with Valves.—Let us now consider the use of valves to complete the chromatic scale on a brass instrument. Such valves act by putting in use additional lengths of tubing and so flatten the pitch (see *Fig.* 14). (This may be contrasted with the method of side openings on the wood-wind instruments which shorten the portion of tube in use and so sharpen the pitch.) For the French horn the longest gap which it is necessary to bridge is that between the second and third open notes *c* and *g*, a musical fifth apart. We thus need six different lengths of tubing to derive, from the upper note *g*, the notes *f*♯, *f*, *e*, *d*♯, *d*, and *c*♯. Then, if this longest gap is bridged, all smaller gaps are bridged also and with something to spare. Further, the valves will be available to extend the scale down from *c*, should that be required. The six lengths of tubing needed are supplied by three valves, of which the middle or second valve adds sufficient length to depress the pitch a semitone, the first valve alone flattens the pitch a tone, while the third valve alone flattens the pitch about a tone and a half. Then the combinations, second and third, give a flattening of about two tones, the first and third two-and-a-half, while all three together flatten the pitch about three tones. It may be noted that the first and second valves are generally used together instead of the third alone. The use of the valves, and the notes produced on a horn (or other valved instrument) in *c*, are shown in Table VIII., together with the order of the partial or open note in use on the total length of tubing.

Any notes above the top C shown would be difficult to obtain and those below the bottom C shown in braces are difficult also.[1]

It is easily seen from the table below that in the higher parts of the compass alternative fingerings are possible. These would be indicated by writing notes in the lines so as to fill up all the blanks, taking care to descend a semitone in each shift to the right. It may be noticed also that two valves (the first and second) would give the chromatic scale over the upper part of the compass. When a longer crook is fitted to the French horn

Fig. 14.—French Horn.

Table VIII

Scale on French Horn in C

Order of Open Notes in Use.	Fingering and Notes.						
	Open 0.	2	1	1 2 (or 3)	2 3	1 3	1 2 3
16	C	..	..	..	..	..	..
15	B♮	..	..	..	..	..	..
14	B♭	A	G♯	..	..	..	..
12	G	F♯	F	..	..	..	..
10	E	D♯	..	..	..	..	..
9	D	C♯	..	..	..	..	..
8	C	B♮	B♭	A	G♯	..	..
6	G	F♯	F	..	..	..	..
5	E	D♯	D	C♯	..	..	..
4	C	B♮	B♭	A	G♯	..	..
3	G	F♯	F	E	D♯	D	C♯
2	C	{B♮	B♭	A	G♯	G	F♯}

[1] On the cornet and some other valved instruments no note higher than the C, which is the eighth open note, is usually employed, this limit being indicated by the dotted line. On the cornet, however, the low notes in braces are all readily obtained.

each of the valves needs more tubing also. They are accordingly fitted with slides that are drawn out to suit.

§ (39) Faulty Intonation of Ordinary Valves.—Let us now inquire if the valves can be tuned so as to be strictly correct for use alone and in combination. Suppose that only the equal temperament is aimed at, so that the problem is simplified as far as possible. Even then the theoretical requirements cannot be met by ordinary valves. Perhaps the simplest way to see this is to take some concrete case as follows: Let the first and third valves give the depression of a musical fourth correctly, *i.e.* the length of the tubing is increased in the ratio 3 : 4. (The tempered intonation is practically equal to the just for this interval.) Then, if the second valve used alone depressed the pitch by an equal-tempered semitone, it will need to be $\frac{4}{3}$ of its previous length to depress it a semitone when used along with the other two valves. If therefore the second valve is not altered when in combination with the others it will depress the pitch by $\frac{3}{4}$ of a semitone only; thus leaving the note in question *one-fourth of a semitone sharp.* Hence, with three ordinary valves, some compromise is inevitable. Thus, as exactness throughout the compass is impossible it has to be considered on what notes mistuning will be least objectionable. A glance at Table VIII., § (38), shows that the first and second valves are more used than the third, which is scarcely needed in the upper range. Also the three valves together are needed only for one or two notes near the bottom of the compass. Further, it must be remarked that the resonance of the instruments under consideration is much more *spread* in the lower part of the compass than in the upper. That is to say, it is quite easy to blow the lower notes a little sharper or flatter than their natural pitches, but difficult to do so with the upper notes. Indeed, other things being equal, the difficulty of blowing a note sharper or flatter than its true pitch is proportional to its frequency.[1] All these considerations point to the desirability of keeping the first and second valves, used alone or in combination, as true as possible and allowing errors to accumulate on the combination of all three. But to minimise the sharpness in this last case the third valve slide may be made fairly long. Let us now illustrate this by a numerical example. Suppose some cornet has an approximate length of 50 inches of tubing from mouthpiece to bell when no valves are in use, and, for simplicity, let us ignore any modifications in lengths that may arise from the fact that through the valves the tubing is parallel but is tapering both before and after, then the lengths to the nearest tenth of an inch required for the various notes in equal temperament are as shown in Table IX.

[1] See *Phil. Mag.*, July 1913.

It is to be noted that the *nominal* note C is really the open note natural to the instrument. Next below are given the lengths that might be added by the ordinary valves, the total lengths of tubing so obtained, and the errors in length and in fractions of a semitone. It is seen that the C♯ and F♯ are about a quarter of a semitone sharp, and that the D♯ and G♯ are about one-seventh of a semitone flat. Also if the third valve be used instead of the combination of one and two, we have the E and A about a quarter of a semitone flat.

§ (40) Equal Temperament with Special Valves.—Let us now notice the principle of certain special valves with which it is possible to attain equal temperament. This improvement has been introduced by two firms, the details of construction being different. In each case, however, the main principle is the same, viz. that when the *third* valve is used in combination with either of the others (or both of them) some *extra* length of tubing shall come into use over that in use with the *first and second* valves alone or together. The effects of this arrangement are shown in the lower part of Table IX., by which it is seen that the lengths shown for the first and second valves are greater when the third is in use than when it is not. The results are seen to be errors so slight as to be practically negligible. Hence, musically speaking, the instrument is true to the equally tempered chromatic scale throughout, provided the open notes are so tuned. Messrs. Boosey & Co., of London, obtain this advantage by arranging that certain extra compensating knuckles of tubing shall come into use in connection with the first and second valves when, and when only, the third valve is depressed. They called their device "compensating pistons."

Messrs. Besson & Co., Ltd., of London, have a pair of separate knuckles of tubing for the first valve, and a pair for the second also. One of each pair is called the *short* slide and comes into play when that valve is used *without* the *third* valve. The other of each pair is the *long* slide and comes into play only when the *third* valve is depressed as well as the valve in question. The arrangement of these pairs of slides (long and short below and above) is clearly seen from *Fig.* 15, showing one of Besson's cornets with their "enharmonic" valves. Each of the five valve slides may be drawn for tuning so as to reach the best adjustment. It seemed desirable to show these valves at the outset in connection with the cornet, but it must be remarked that their use is more desirable in the

TABLE IX

VALVES ON CORNET

True total lengths needed for equal temperament		Inches. 50	Inches. 53·0	Inches. 56·1	Inches. 59·5	Inches. 63·0	Inches. 66·7	Inches. 70·7
Lower notes on instrument		G C	F♯ B♮	F B♭	E A	D♯ G♯	D G	C♯ F♯
Ordinary Valves.	Valves and the lengths added 2	..	3·2	..	3·2	3·2	..	3·2
	1	..	..	6·3	and 6·3	..	6·3	6·3
	3	..	..	..	(10·3)	10·3	10·3	10·3
	Total lengths obtained	50	53·2	56·3	59·5 (60·3)	63·5	66·6	69·8
	Errors in length	0	+0·2	+0·2	0 (+0·8)	+0·5	+0·1	−0·9
	Errors in fractions of a semitone	0	1/15	1/15	0 (1/4♭)	1/7♭	1/37	0·23♯
Special Valves.	Valves and the lengths added 2	..	3·15	..	3·15	3·75	..	3·75
	1	..	..	6·25	and 6·25 or	..	7·4	7·4
	3	..	..	..	9·4	9·4	9·4	9·4
	Total lengths obtained	50	53·15	56·25	59·4	63·15	66·8	70·55
	Errors in length	0	+0·15	+0·15	−0·1	+0·15	+0·1	−0·15
	Errors in fraction of a semitone	0	1/20	1/20	1/34	1/23♭	1/37	1/27♯

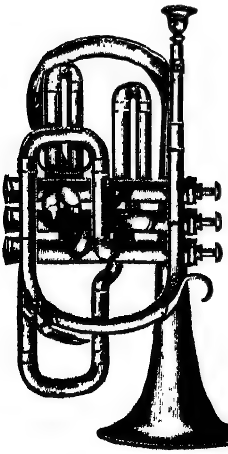


FIG. 15.—Cornet with Special Valves.

euphonion and bombardon. These will be dealt with later. The much greater masses of air in these bass instruments make it much harder for the performer to correct any faulty intonation by blowing so as to force the notes in tune. Whereas, with smaller instruments like the cornet and trumpet, professional players often disdain all complications, since they are so expert at forcing each note into tune.

§ (41) THE BACH TRUMPET.—This trumpet is quite straight and allows the light to pass through it when the valves are not depressed. It is usually made in E♭ and D, the change being made by using the E♭ or the D middle piece between the valves and the bell. (See *Fig.* 16.)

As in the case of the French horn, this trumpet is made with two or three valves according to the musical passages for which it is required. Thus in Handel's *Messiah*, for the trumpet obbligato in "The trumpet shall sound," the solo player generally uses an instrument with two valves only, the third valve not being required. Indeed this part was originally written for a valveless trumpet

of double length and therefore an octave lower. Then the high F and F♯ had to be blown into tune from the eleventh open note, which really lies between them in pitch. This trumpet has a tone of great nobility and brilliance, but yet is susceptible of piano passages and is then comparatively mellow. It reaches easily the highest notes allotted to the orchestral brass. It is tuned by the particular setting of the middle piece between the valve portion and the bell, a certain latitude being available at each end. To put the instrument in the low pitch (or "new Philharmonic") a small *bit* of tubing about an inch and a quarter long is inserted next to the mouthpiece. The valve slides are all drawn the correct distances when this bit is in use or when the trumpet is put in D.

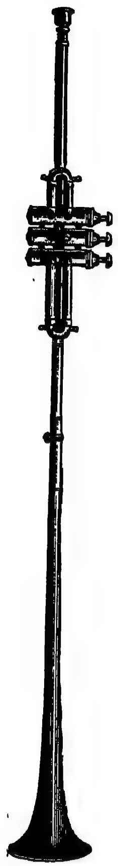


FIG. 16.
Bach Trumpet.

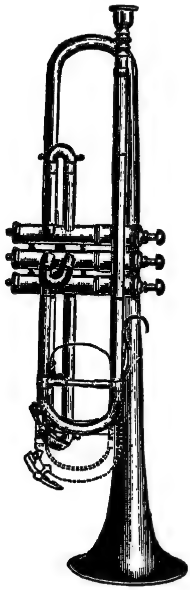


FIG. 17.
Trumpet in B♭ and A.

§ (42) TRUMPETS IN B♭, F, ETC.—The trumpets in ordinary use have the tubing coiled up in the fashion so familiar and as shown in *Fig.* 17. This model changes easily from B♭ to A♮ by drawing the slide as indicated lightly at the right. This kind of trumpet is now made in C also and is much in use in orchestras. *Fig.* 18 shows the symphony trumpet in C with rapid rotating change to B♮ made by C. G. Conn, Ltd., of Elkhart.

Trumpets in F and E♭ of the same model as the above, but without any rapid change, are often used in orchestras for the lower trumpet parts. They can be set to E♭ by the insertion of a crook as shown in *Fig.* 19.

Trumpets in B♭ are also made without valves for fanfares, etc. Their scale is then restricted to the notes of the harmonic series of relative frequencies [1], 2, 3, 4, 5, 6, [7], 8. The first and seventh notes are not generally used. (See *Fig.* 35, § (52) (vi.).) Bugles in B♭ are similar as to the absence of valves and consequent restriction of scale, but are more gently tapering from the bell and so have a mellower, less brilliant tone. (See *Fig.* 34, § (52) (vi.).)

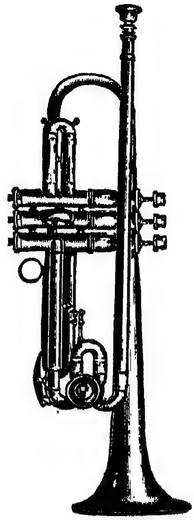


FIG. 18.
Trumpet in C and B♮.

§ (43) TROMBONES.—We now consider a family of brass instruments in which no restriction is placed on the intonation possible. Thus it may be equal temperament or just anything that is wished and the performer has the skill to attain. This is owing to the fact that the mechanism for the scale consists of a U-slide which may be drawn out so as to flatten the pitch continuously by any desired amount to six semitones. When the slide is closed it is said to be in the first position. By extending it to the second position the pitch is flattened a semitone, on reaching the third position the pitch is a whole tone lower, and so on to the full extension of the slide, which is the seventh position and involves a flattening of six semitones. On comparing this arrange-

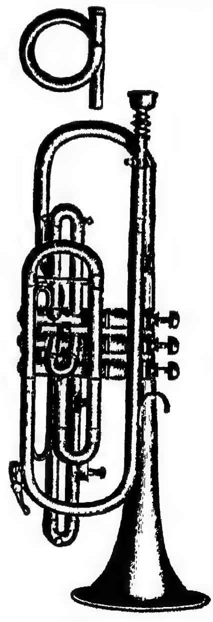


FIG. 19.
Trumpet in F and E♭

ment with the system of three valves it is seen that the first position of the slide corresponds to the open notes on the valved instrument, the second position to the depression of the second valve alone, the third position to the use of the first valve alone, and so on, the seventh position corresponding to the use of all three valves in combination. The full comparison may be shown thus:

Positions of slide	1st.	2nd.	3rd.	4th.	5th.	6th.	7th.	
Flattening involved	0	1	2	3	4	5	6	semi-tones.
Corresponding valves	0	2	1	1 2	2 3	1 3	1 2 3	

Slide trombones are made of various pitches called alto, tenor, and bass trombones; these are in E♭, B♭, and G respectively. A tenor trombone is shown in *Fig.* 20. At the top end is seen the tuning slide, and it may be noticed that its plane is at right angles to that of the long slide used for the scale, which extends downwards from the mouthpiece. This setting of the two bends at right angles is a peculiarity of slide trombones, and distinguishes them from all other instruments, in which the coiling of the tubing is almost parallel to one plane. Although the slide of the trombone allows of perfect intonation, it calls for additional skill on the part of the performer to take all the positions with readiness and right adjustment. Further, it calls for great care on the part of the makers as to the right proportioning of the taper and curvature of the tubing. In any brass instrument there is much difficulty in obtaining a tube of such shape as to give the harmonic series of open notes in tune down to the second. This difficulty is enhanced when the taper is interrupted by a parallel portion to act as a tuning slide or a set of valves. Now in the case of the three ordinary valves there must be a length of parallel tubing provided sufficient to flatten the pitch three tones. The part of this length in use varies from nothing to the full amount. But in the slide trombone an additional length of parallel tubing sufficient to flatten the pitch three tones must be drawn out for the seventh position, and the same length of parallel portion must be present for the slide to close upon into the first position. Hence in the slide trombone the length of parallel tubing provided is *double* that required for three tones, the amount in use varying from the half to all of it. Again, in the slide of the trombone we have a slight change of diameter inevitable on passing from the parallel fixed portions to the movable portions of tubing which slide upon them. To minimise complications and avoid a "break" upon certain notes this difference of diameters must be as small as possible.

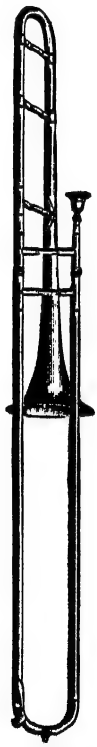


FIG. 20.
Tenor Trombone in B♭.

The trombones are capable of great power and have a rich and brilliant tone. In the hands of sympathetic performers they are also capable of *pianissimo* effects of great beauty and value. The somewhat uncouth mechanism for the scale forbids their use in quick florid passages, especially in the lower parts of the scale where the extreme positions of the slide are needed. They shine rather in slow movements of a certain stately grandeur. It may be noticed in conclusion that the possession of the slide confers the power to produce the grace called *portamento*, or the continuous gliding of the pitch from one note to another. This power in its complete fulness is otherwise possessed only by the instruments of the violin family and by the human voice.

A slide trumpet is occasionally used for oratorios and other purposes. It is very little different from the trombone in essential qualities, so is included here (see *Fig.* 21). By the use of crooks it can be put in various keys.

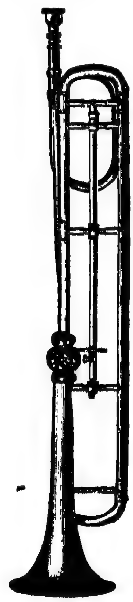


FIG. 21.
Slide Trumpet.

§ (44) VALVED INSTRUMENTS OF BRASS BAND.—We now notice the instruments which form a complete family of similar units and constitute the brass band proper, to which may be added trombones and trumpets. All the various elements of this family have valves, usually three only, but some of the

bass ones occasionally have four or even five valves. Beginning at the upper part of the compass and passing down to the bass, we have in order the following chief instruments: E♭ soprano, B♭ cornet, E♭ tenor horn, B♭

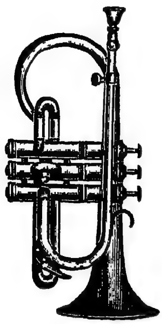


FIG. 22.—E♭ Soprano.

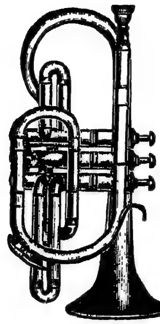


FIG. 23.—B♭ Cornet.

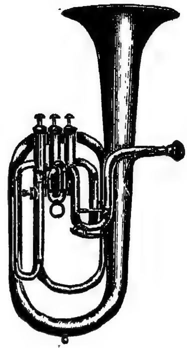


FIG. 24.—E♭ Tenor Horn.

euphonion, E♭ bombardon, BB♭ monster bombardon. These are shown in *Figs.* 22 to 27. Others of secondary importance are B♭ fluegel horn, same pitch as the cornet; B♭ baritone, same pitch as the euphonion but of smaller tubing; B♭ bombardon, same pitch as BB♭

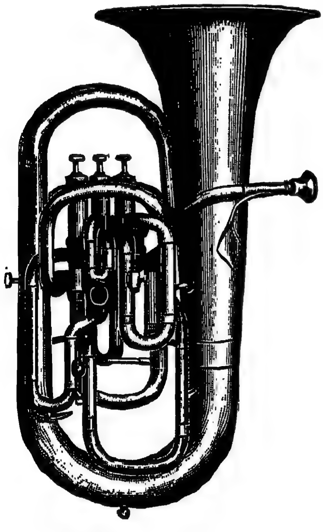


FIG. 25.—B♭ Euphonion.

monster but of smaller tubing. Thus, glancing at the list, we see that the tenor horn is an octave below the soprano, while the bombardon in E♭ is an octave below the tenor horn. Again the euphonion is an octave below the cornet, while the B♭ bombardon is an octave below the euphonion. (The orchestral tuba in C is a note above the B♭ bombardon.)

Of the main instruments, the first four, soprano, cornet, tenor horn, and euphonion, may be compared to the four varieties of human voice, soprano, contralto, tenor, and bass. The bombardons differ from the euphonion more in getting greater power on

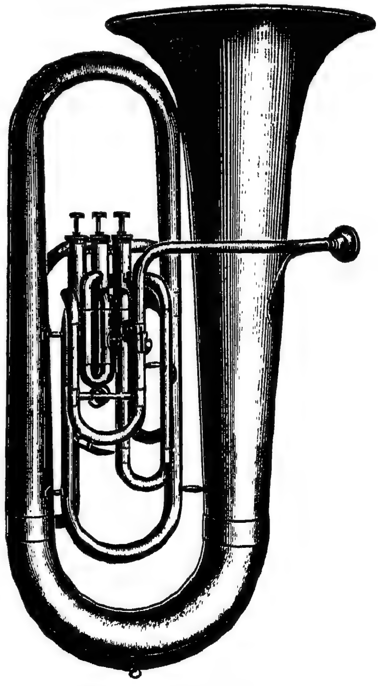


FIG. 26.—E♭ Bombardon.

the low notes than in going so much lower. This is due to the fact that the euphonion is often provided with a fourth valve to enable it to take the full chromatic scale down to its

pedal B♭. This fourth valve lowers the pitch about a musical fourth, so by it, in conjunction with the others, the player can bridge the gap of an octave between the

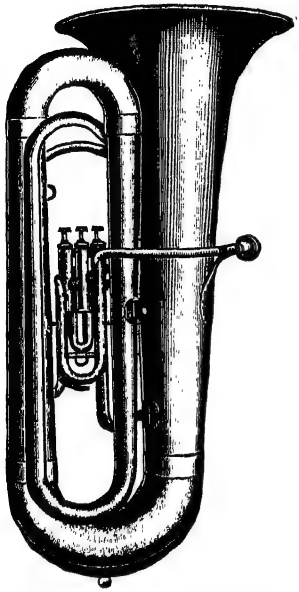


FIG. 27.—BB♭ Monster Bombardon.

second open note and the pedal or first of the same name. This device is occasionally applied, however, to the E♭ bombardon also. In either case the fingering is as follows, the notes being written C to C as is usual for any valve instruments playing from the treble clef:

Notes in pedal octave	C.	B.	B♭.	A.	G♯.	G.	F♯	F.	E.	D♯.	D.	C♯.	C=pedal.
Fingering of four valves	0	2	1	3	2 3	1 3	1 2 3	1 4	3 4	2 3 4	1 3 4	1 2 3 4	0

In these bass instruments the compensated pistons or enharmonic valves are specially important. For the mass of air to be set in vibration is very much greater than in the case of higher and smaller instruments. Consequently it is harder for the player to elicit a full and powerful tone unless the resonance is at its best, because no forcing to a somewhat different pitch is needed. It may also be seen that with uncompensated valves the defect would be emphasised by the addition of a fourth valve to be used with them down to the pedal.

Of the instruments now under review, each has a somewhat different voice and utterance (as well as compass), and each has its appointed place to fill. The cornet is specially prized as a solo instrument, and next to it the euphonion, the two being roughly comparable to a contralto and a bass soloist. By the device of double-tonguing (borrowed originally from the flautist) very rapid iterations on the cornet are quite easy. Also the attack may be varied from one of almost ideal smoothness to the other extreme of declamatory abruptness, the tone very varied to match. Staccato and legato playing are equally easy, shakes and turns (though not quite so fine as on the strings or wood-wind) can be readily executed, and just a slight suggestion of *portamento* is possible. A wide range of power from double *forte* to *pianissimo* is at the disposal of the competent performer, either over a succession of detached notes or in the *crescendo* and *decrescendo* of a sustained note. Further, the grace called *tremolo* is possible on the cornet and may be easily indulged in too freely. This embellishment appears to consist of an alternate waxing and waning of the sound a few times per second, produced by vibrating the instrument in the hand. But the cornet may be and often is vulgarised, as it lends itself so easily to music that is frivolous and ephemeral. Most of the above powers belong as well to the trumpet as to the cornet, but the trumpet, having a more noble and brilliant tone (and often associated with wood-wind and strings), is usually reserved for loftier utterances and is concerned rather with the clash of battle or the pomp of regal appearances than with the more commonplace occasions of life and the sentiments appropriate to them.

Comparing for a moment the piano, organ, bowed strings, and wind instruments, we see that the first lacks sustaining power, the second lacks accent, while the rest have both. In compensation, the first two are harmony instruments in themselves, while the rest are only melodic components of a concerted harmony.

§ (45) HUMAN VOICES.—In many respects, to the human voice must be conceded the highest place among musical instruments. For the variety and depth of expression which can be concentrated into a single note, it stands unrivalled. Though the compass of any one voice may be less than that of most instruments, the four varieties taken together range over and exceed the limits of the treble and bass staves. Among the women's and men's voices and the boy sopranos we have also a delightful variety of tone quality and aptitude for music of distinctly different characters. But the chief characteristic of the human voice is its unique power of so modifying the quality of the sustained sounds

and so specialising their start and cessation as to give utterance to vowels and consonants. That is to say, the human voice alone can render the words as well as the music. For the vowels consist of a very special quality of sustained tone which may be called *vowel quality*. Of the consonants some are produced by specially rapid and irregular vibrations preceding or following an intermediate vowel. Some other so-called consonants can be sustained as long as any recognised vowel and have scientifically the right to be classed as vowels. The letter *s* is a case in point.

This is not the place to enter into any minute anatomical examination of the human organs of song and speech. A very cursory notice of a few salient features must suffice here. Adhering then to our usual subdivision of the essential parts of a musical instrument, we may, for the human voice, state briefly as follows. The *exciter* is the breath from the lungs *via* the windpipe. The *vibrator* of definite pitch is the pair of vocal cords (or ligaments), which form a partial obstruction across the larynx (or voice-box) and leave only the vocal chink or slit between them for the passage of air. The *resonator* is triple, consisting of the cavities of the pharynx (or upper part of the throat), the mouth, and the nose. These are modified by the tonsils, the soft palate, the tongue, and the lips. The *mechanism* for the scale consists of the set of muscles which control the position, tension, and therefore also the vibrations of the vocal cords. The vocal cords in a state of rest are about three-quarters of an inch long in men, and about half an inch long in women. When producing a musical sound the vocal chink (or slit between the vocal cords) is alternately closed and opened at the required frequency, depending on the tension of the vocal cords, their length, mass, and loading. The air-blast is, in consequence, checked and allowed to pass alternately and at the right frequency. Thus the loudness and pitch of a note depend upon the amplitude of vibration and frequency of the vocal cords, and consequently upon the muscles which supply the blast and set the vocal cords in the precise state required. It may be remarked that the resonator, having soft walls, is unable to modify the pitch fixed by the vocal cords. Upon what does the quality of the tone depend? Clearly, (*a*) upon the exact nature of the vibrations executed by the vocal cords and so impressed on the issuing air, and (*b*) upon the resonant cavities which modify these vibrations, encouraging some partials and relatively discouraging others. That special quality which we recognise as constituting a given vowel appears to be due to the special favouring by resonance of any partials near a given *fixed pitch, whatever the pitch of the prime* may be. Thus, the vowel ōō, as in moon, is characterised by the favouring of those partials whose frequencies are near 175 per second. Accordingly, a musical note whose prime was 87·5 per second with a very loud *second* partial (175) would sound like ōō; so also would a musical note whose prime was 175 per second, if the *prime* itself were very loud compared with any of the upper partials. It is easily seen that the cavities affecting the human voice may be set at will to respond to various pitches and thus give utterance to the corresponding vowel. Further, if a given vowel is sung in notes of different pitches, then the setting of palate, tongue, and lips must be preserved unchanged to keep the same vowel (according to the above statement) while the vocal cords are altered to change the pitch of the prime. And this is a matter of experience, as any one may verify.

§ (46) Comparison of the Voice with other Instruments.—We may now naturally compare the human voice with other musical instruments as to the power left to the performer to control the quality of tone produced and in other respects.

In the case of the piano the quality varies somewhat with loudness, but probably the performer has little or no control over it for any specified degree of loudness, though he often imagines he has. Indeed it seems scarcely possible that he should have such control. For the maker has settled the shape and quality of surface and mass of the hammer and the place on the string to be struck by it, also the sound-board and bridges; and upon all these the quality depends. What choice remains to the performer concerning an individual note played without use of pedals? The speed at which the hammer shall strike the string. This appears to be the sole variable at the disposal of the performer, especially in the case of pianos in which the hammer is free from the mechanism of the action before it strikes. But of course for each note struck there may be a different speed of striking in use and consequently a different quality of tone elicited. Thus an air or certain notes in it may be made distinct from those of the accompaniment by a different degree of loudness and consequent difference in quality of tone, and the power to produce this difference is at the instant disposal of the performer.

In the organ it is the maker again who has controlled the quality of tone in the pipes of each stop. The performer has no power over that and can simply mix as he chooses what is provided for him. But even so, what a vast field is placed at his disposal! Where only ten stops are available over a given compass, if all were suitable for the purpose, no less than 1023 different combinations are

possible. (This is easily seen if we note that there are 10 ways of taking a single stop and 10 more for nine stops, 45 ways of taking two stops and 45 for eight stops, 120 ways for three stops and 120 for seven, 210 ways for four stops and 210 for six stops, 252 ways for five stops, and, finally, only 1 way of taking all ten.) Again, for twenty stops the number of combinations is 1,048,575, or 500 per week for about 42 years! The varieties of tone quality at the disposal of the organist are, however, available only for broad effects, and not for single notes as in the case of the pianist.

In the case of the violin much has been done by the maker and the age and ripening of the instrument, but very much is still left to the performer. He may re-set the sound-post to his liking, he may choose his strings to suit the particular instrument, also his bow, and take care that it is well haired and properly rosined. He may often choose upon which string a certain passage shall be taken. He must always choose the place of bowing, the pressure and the speed, and can vary these at will for each string, for every stop upon it, and for every instant while the note lasts. And all these variables affect the quality of the tone.

In the case of the wood-wind there is the right application of the player's lips to the flute, or to the reed of the oboe or clarinet, and the appropriate pressure and air supply. And these variables must be changed in concert for the different notes and for every variety of light and shade on each.

On the brass instruments there is the right application and tension of the lips to the mouth-piece, and the corresponding suitable pressure of air and speed of supply to be arranged and maintained for each note of the scale and each degree of loudness or delicacy with which it is produced.

But, in playing all the foregoing instruments, there is no possibility open to the performer of appreciably varying the character of the resonance arrangements applicable to each note. That has been settled beforehand by the maker of the instrument.

In these respects the human voice has at the command of the singer resources not offered by any musical instrument fashioned by man. For not only can the human voice be controlled at the very seat of its production, where the breath passes between the vocal cords, but the resonant cavities which modify the original vibrations are all under control also. Hence the final resulting quality may be varied from instant to instant so as to render tones expressive of triumph, menace, or entreaty, or passing from artless gaiety to pathetic wailing.

§ (47) Drums.—As is well known, the exciter of a drum is the drumstick, the vibrator is the circular membrane called the drumhead or skin, the resonator is the chamber between the two drumheads, or, in the case of kettle-drums, the space below the single head. In the case of ordinary drums, bass or side, there is such a jumble of sounds produced that it is extremely difficult to assign any definite preponderating pitch to the resulting effect. In the case of kettle-drums it is quite possible for the trained ear to detect the preponderating pitch and to tune the drums to the note required. In *Fig.* 28 is shown an orchestral kettledrum in which the T-shaped tuning handles are clearly seen. It is customary to have three kettledrums in an orchestra, all under the charge of a single drummer. The drumsticks may be entirely of wood, or have felt heads, or have heads of cork covered with chamois leather. Berlioz expressed a preference for drumsticks with their heads covered with sponge.

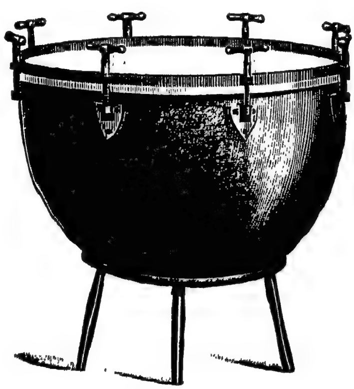


Fig. 28.—Kettledrum.

§ (48) Lateral Vibrations of Bars and Tubes.—If a thin uniform bar of metal, glass, or wood is supported at two places rather more than a fifth of its length from its ends, and then struck or otherwise excited, it will execute its slowest lateral vibrations and sound its corresponding fundamental note. Other higher notes are possible to the bar if supported at two of the places which are nodes for those higher sounds. If the fundamental for such a bar be the note *c*, then the first few of the series are approximately *c*, *f'*♯, *f''*, *d'''*, *a'''*. Expressed more accurately by their relative frequencies these would be 1, 2·756, 5·404, 8·933, and 13·345 respectively. It is seen that the possible partials of this set rise in pitch much quicker than those of strings, open parallel pipes, and open-ended cones closed at their vertices. Indeed, except for the second partial, they rise quicker even than those for a stopped parallel pipe, which

has the series 1, 3, 5, 7, 9, etc. The sounds of these bars afford a good example of inharmonic partials, *i.e.* those whose frequencies are inexpressible by small whole numbers.

Sets of bars supported so as to emit their fundamentals when excited, and arranged to give the chromatic scale over about two octaves, are used occasionally in orchestras under the name of *glockenspiel* (chime). The bars are struck by little mallets or strikers

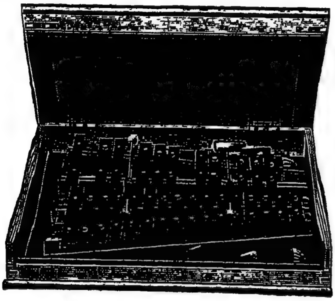


FIG. 29.—Glockenspiel.

(see *Fig.* 29). Bars with both ends free, as here dealt with, are often called *free-free* bars.

Instead of bars laid horizontally and supported at two points, tubes (or bars) may be hung vertically and struck with a mallet or beater (padded or bare) and so imitate bells to some extent. They are usually made in octave sets of 8 tuned to the diatonic scale, or 13 tuned to the chromatic scale. (See *Fig.* 30.) On referring to § (50) it will be seen that the quality of tone of a church bell is different from that of a plain bar or tube, since its partials form a distinctly different series.

§ (49) TUNING FORKS.—As a standard of pitch for ready reference a tuning fork is very convenient as it has no need of supports at special places like a free-free bar. Each prong of the fork is very like what is called a *fixed-free* bar, being practically fixed in the massive part near the shank where the prongs unite, but free at the other end. A single fixed-free bar would be very unsuitable as a standard vibrator since it would be unbalanced and therefore need a very massive base to fix one end. The tuning fork always vibrates with its prongs both out or both in together, and so balances itself and consequently needs very little support. It should be noted that a fork lightly and carefully bowed at or near the ends of the prongs emits practically its prime partial only. Its second partial is very high, having about 6·6 times the frequency of the prime. This is obtained when a fork is struck on a bare table or other hard surface. The prime is much fuller when the fork is mounted on a suitable resonance box.

§ (50) BELLS.—As we have already seen, the musical *string*, whether of catgut or metal, is very simple in its behaviour and yields the full series of harmonic partials. A membrane, though flexible like the string, is two-dimensional, having length and breadth instead of being almost confined to length only like the string. Hence its vibrations are much more complicated than those of a string. We may approach the vibrations of a bell by way of those of a straight bar, which we have seen consist of a set of inharmonic partials. When the straight bar is replaced by a bent one in the form of a fork with special mass and shank at the bend, the vibrations are still more complicated. If we now pass to the case of a bell, whose sides are like a number of prongs all united into one of circular form, we have far more complication still, even if the sides and top were uniformly thin. If the top and the different levels of the side have all varieties of thickness, then the state of things is quite beyond rigorous mathematical treatment. We can only derive

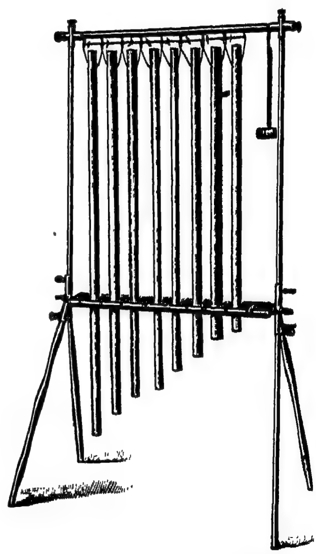


FIG. 30.—Tubular Chimes.

some guidance and light from a combination of theoretical speculation and experimental tests. Thus we might expect the bell to show *nodal meridians*, or lines where an axial plane cuts the surface, and along which there is no radial displacement. Again, we might expect there would be nodal circles where planes perpendicular to the axis cut the surface. Both these surmises are verified by experimental tests. It should be noted that a

node for radial motion may be an antinode for tangential motion. Bell-founders recognise five chief tones or partials in the sound of a church bell. Beginning from the highest these are called the nominal, quint (or fifth), tierce (or third), fundamental, and hum-note. The nominal, fundamental, and hum-note are spaced about an octave apart. The other two approximately fill in the fifth and *minor*-third in the upper octave. At one time none of the five were in any exact relation that could be expressed by small whole numbers, *i.e.* all the partials were inharmonic. But, by turning or boring at certain regions of the bell, it was afterwards found possible to get the three main notes in the true octave relation or nearly so. It has now been asserted that one firm of bell-founders is able to tune *each* of the five tones of a church bell true to a single vibration per second.

§ (51) FOG-HORNS.—For the warning and guidance of mariners fog-horns are erected on capes and other important points, and serve the same purpose during a fog that a lighthouse serves during the night. Small ones may be worked by hand power, but large ones are driven mechanically by an engine or motor of some kind. The passage of air in the throat of the apparatus may be alternately interrupted and freed by a revolving disc or cylinder with holes or slots which pass similar openings in a corresponding fixed plate or cylinder. Such an arrangement is called a siren. The sound thus originated is reinforced and directed to its objective by a large horn. This may be twenty feet long and taper from a few[1] inches diameter at the throat to five or six feet across at the mouth of the bell. If (as is usual) the sound is required to spread over a large distance horizontally and a less distance vertically, then the mouth of the bell is made elliptical with the major axis vertical and the minor axis horizontal. For the less any diameter of the bell is the better does the sound spread in the directions of that diameter. This is an example of the *diffraction* of sound, a phenomenon more frequently met with and studied in connection with light.

§ (52) FREQUENCIES OF SOUND PRODUCERS CALCULATED. We now notice how the frequency of a simple sound producer may be calculated so that arrangements may be made accordingly in any actual case.

(i.) *Strings.*—In the case of strings displaced transversely we may note first that the speed[2] of propagation of such a disturbance is given by $v=\sqrt{(F/m)}$, where F is the stretching force in dynamic units and m is the mass of the string per unit length. Thus, if we want v in cm. per sec., we should take F in dynes and m in gm. per cm. If, on the other hand, F is expressed in poundals and m in pounds per foot, v will be obtained in feet per sec.

Suppose now a harp string be plucked near one end, then the disturbance passes along to the far end, is there reflected with a *reversal* of the displacement, comes back to the end where it first occurred, and is there a second time reflected with a reversal. Hence after passing over twice the length of the string the original state of things is almost repeated. Thus, the time in question must be the period of vibration. Or the reciprocal of that must be the frequency of the string. But this will be the number of double traverses accomplished in one second, and therefore equal to the speed of propagation divided by twice the length of the string. Thus, for the frequency of the fundamental (or prime) of the string, we have the expression

$$N_1=\frac{v}{2L}=\frac{\sqrt{(F/m)}}{2L}, \quad . \quad . \quad (1)$$

where N_1 is the frequency of the prime for a string of length L. The violin family falls under this class. So the pitch is raised by stopping a string on the finger-board and thereby shortening its vibrating length.

For the other possible partials of the string, we imagine it vibrating in two segments with a node at the middle, in three segments with two nodes at the points of trisection, and so forth. Thus the virtual lengths for these become half, third, and so forth of the whole length. Hence the upper partials have frequencies double, treble, etc., that of the prime. Or

$$N_1=\frac{N_2}{2}=\frac{N_3}{3}=\frac{N_4}{4}, \quad . \quad . \quad (2)$$

where N_2 is the frequency of the second partial, etc. On this principle *harmonics* are obtained on the harp, violin, etc.

(ii.) *Open Pipes.*—The case for parallel pipes open at both ends is very similar to that of strings. We are concerned with the speed of propagation in the air[1] or other gas filling the pipe, and with the number of traverses of its length L before the original state of things is approximately repeated. (See § (33) (i.).) Thus, for the open pipe, since at each end a compression is reflected as a rarefaction and *vice versa*, two traverses complete the cycle and constitute the period. Or the frequency of the prime is the speed divided by the

[1] Sometimes surprise is expressed that sufficient range can be obtained with so small a throat. No such difficulty need be felt when it is remembered that the mouthpiece of a cornet narrows to one-fifth of an inch in diameter and that its sounds may be heard with ease half a mile away.

[2] See "Strings, Vibrations of"; also Rayleigh, *Sound*, i. chap. vi.

[1] See "Vibrations of Air in a Tube."

double length of the pipe. Hence we may write

$$N_1=\frac{v}{2L}=\frac{\sqrt{(kP/D)}}{2L}=\frac{v_0\sqrt{(1+(t^\circ/273))}}{2L}, \quad (3)$$

where N_1 is the frequency of the prime, L the length of the pipe, v the speed of sound in the gas, P and D its pressure and density at temperature t° C, k the ratio of its specific heats, and v_0 the speed of sound in the gas at 0° C.

By subdivision of the pipe into segments we see that quicker vibrations can occur, and that the conditions for these segments are (a) antinodes at ends, with (b) freedom to have either nodes or antinodes at other places in the pipe. Hence the pipe may subdivide in 2, 3, 4, etc., segments. (See *Fig.* 5, § (33).) Thus we have for the frequencies N_2, N_3, N_4, etc., of the upper partials

$$N_1=\frac{N_2}{2}=\frac{N_3}{3}=\frac{N_4}{4}, \text{ etc.} \quad . \quad . \quad (4)$$

(iii.) *Corrections for Mouth and Open End.*—We have hitherto supposed that the reflection occurs exactly at the end of the pipe. In reality there is an end correction needed,[1] as though the pipe were a little longer, or as though the vibrations extended a little beyond the end before reflection occurred. For a cylindrical pipe without any flange this correction is three-tenths of the diameter. Further, if the fully open end needs a correction, as though it were not open enough, much more must there be a correction for the mouth, which is far less open. This correction is of the order 1·36 times the diameter. Thus the two together make about five-thirds the diameter. Hence, if we denote by e and m the end and mouth correction respectively, (3) and (4) may be rewritten thus,

$$N_1=\frac{v_0\sqrt{(1+(t^\circ/273))}}{2(L+m+e)} \quad . \quad . \quad (5)$$

$$N_1=\frac{N_2}{2}=\frac{N_3}{3}=\frac{N_4}{4}, \text{ etc.} \quad . \quad . \quad (6)$$

The flute falls under this category. So the pitch is raised by opening a side hole and thereby shortening the vibrating length in use. But it is noteworthy that the end correction e will be much greater for a side hole than for the open end of the flute. And the smaller the hole the greater the correction.

(iv.) *Stopped Pipes.*—As previously shown (see § (33) (ii.)) the disturbance passes four lengths of a stopped pipe before the original state of things is repeated. Hence for the frequency of the prime we may write

$$N_1=\frac{v_0\sqrt{(1+(t^\circ/273))}}{4(L+m)}, \quad . \quad . \quad (7)$$

where v_0 is the speed of sound in the gas at 0° C., t is the temperature of the gas in the pipe, L is its length, and m the correction for the mouth.

Further, as we have seen (§ (33) (ii.)) that the second partial has three times the frequency of the first, the third partial five times the prime frequency and so on, we have

$$N_1=\frac{N_2}{3}=\frac{N_3}{5}=\frac{N_4}{7}, \text{ etc.} \quad . \quad . \quad (8)$$

The clarinet falls approximately under this category, hence its special tone quality and its overblowing the twelfth. It is also seen that the opening of its side holes shortens the vibrating length and so raises the pitch.

(v.) *Conical Pipes.*—Let us now consider the case of a conical pipe closed at the vertex and open at the base. Then it is found that there is an antinode at the open end and that the other antinodes occur equidistantly just as if the whole pipe were parallel and open at both ends. The nodes, on the other hand, are displaced towards the apex from the intervening equidistant positions they would occupy in a parallel pipe. Moreover, those antinodes near the open end are displaced a little, the others more and more till the last is always displaced right up to the vertex. (See *Fig.* 31.) Thus, with due correction for the open end, the possible frequencies for this conical pipe are given by

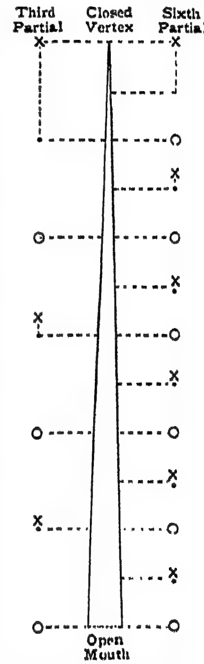


Fig. 31. Conical Pipe.

$$N_1=\frac{v_0\sqrt{(1+(t^\circ/273))}}{2(L+e)}=\frac{N_2}{2}=\frac{N_3}{3}=\frac{N_n}{n}, \text{ etc.,} \quad (9)$$

where N_1, N_2, etc., are frequencies of prime and partial, v_0 is speed of sound in the gas in use at 0° C., L is the actual length of the pipe in use, and e the end correction. In the figure the antinodes are shown by circles, the nodes by crosses, and their undisplaced position by dots. The pipe is shown speaking its third and its sixth partials.

This case includes (approximately) the oboe and bassoon. It should be observed again that the end correction e will be much

[1] See Rayleigh, *Sound*, ii. 201; also *ibid.* chap. xii.

more for a small side hole than for the termination of the pipe at the bell mouth. It is easily seen that the mouthpiece and bell involve a slight departure from the simple form of the ideal cone.

(vi.) *Hyperbolical Pipes with Mouthpieces.*—The case of all the brass instruments with cupped mouthpieces falls under this class, which accordingly merits special attention. It may seem that the simple cone and the larger brass instruments have very little in common, but the wide gap between them may be bridged by a number of quite small steps. Take first the tandem horn shown in *Fig.* 32, and next the drag horn shown in *Fig.* 33.

FIG. 32.
Tandem Horn.

FIG. 33.
Drag Horn.

The first of these departs but slightly from the simple cone, the second rather more. But neither of them disturbs materially the sequence of the partials possible to the pipe. This is on account of the *hyperbolical curvature* of the sides, which is purposely introduced in order to counteract the disturbing effects of the bell and the mouthpiece. The great length of such straight tubes is often very inconvenient. Therefore they are often coiled up into a compact form. Thus, the tandem horn gives place to the bugle (see *Fig.* 34), and the drag horn to the cavalry trumpet (see *Fig.* 35).

When the ideal cone is departed from by introduction of bell and mouthpiece, it is difficult to so model the pipe as to retain *all* the partials in their strict harmonic relation. The practical escape from this difficulty for musical purposes is highly interesting. Such a curvature is adopted as will keep all partials in relative tune *except the pedal*, or prime of the whole series. A mouthpiece is adopted which is usually *too small* to allow of the easy production of the pedal. Then the others are all musically available, and the largest gap to be bridged by valves or slide is a *fifth* (between open notes 3 and 2) instead of an octave (between open notes 2 and 1). The resultant set of partials for some forms of tubing involves a sharpening of the pedal from its pitch as in the cone, but more often we find in use a form of tubing involving a flattening of the pedal.

FIG. 34.
The Duty Bugle.

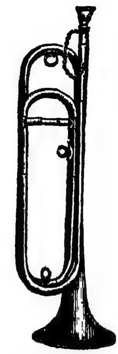


FIG. 35.
The Cavalry Trumpet.

The two cases are contrasted with the simple cone in the following approximate statements of frequencies.

Trumpet and Cornet .	$\frac{9}{8}$, 2, 3, 4, 5, 6, 7, 8, etc.
Simple Cone . . .	1, 2, 3, 4, 5, 6, 7, 8, etc.
Euphonion and Bombardon	$\frac{7}{8}$, 2, 3, 4, 5, 6, 7, 8, etc.

If, therefore, we seek the length for a brass instrument of a given pitch, it is easily seen that there is the choice to be made between calculating for the mistuned pedal or for the other notes musically in tune. Again, if the pitch is to be right at a given temperature of the room, there is the further uncertainty as to the actual temperature (and moisture) of the player's breath in the instrument when in such a room. It is difficult to settle any of these points with precision. In the meantime it is instructive to work out a concrete case for a simple cone and compare the result with an actual instrument, say the B♭ cornet, whose length has probably been fixed by experience. The military pitch of the treble B♭ is 479·3 vibrations per second at 60° F. $= 15\frac{5}{9}$° C. Thus for the length L of the cone with end correction e we have

$$L + e = \frac{1089 + (2 \times 15\frac{5}{9})}{2 \times (479 \cdot 3 \div 4)} = \frac{1120\frac{1}{9}}{239 \cdot 6} = 4 \cdot 675 \text{ ft.} \quad (10)$$

If the diameter of the base is 4·5 in., we have

$$e = 1 \cdot 35 \text{ in.}$$

Hence

$$L = 4 \cdot 675 \times 12 - 1 \cdot 35 = 56 \cdot 1 - 1 \cdot 3 = 54 \cdot 8 \text{ in.}$$

The length of the tubing of a B♭ cornet by Boosey & Co. measures 54·5 in. when in the military pitch. It thus appears that the various departures of this case from the simplicity of the cone are almost compensatory. As to what some of the separate departures are we may note the following.

The player's breath has a temperature well over 80° F., as may be found by breathing on a thermometer. Suppose the breath leaves the cornet at about 64° F. (as a simple test showed), then the mean temperature in the instrument may be taken at say 77° F. or 25° C. Further, the moisture due to the breath might be accountable for an extra 5 ft. per second in the speed of sound. Then the speed of sound in the cornet might be $1089+(2\times 25)+5=1144$ ft./sec. Again, taking the pedal as expressed by $\frac{5}{8}$ instead of unity, its frequency would be $\frac{5}{8}$ of $(479\div 4)=100$ per sec. nearly. Then, finding the length as though it were not changed by the hyperbolic shape, we should have

$$L+e=\frac{1144}{2\times 100}=5{\cdot}72 \text{ ft.}=68{\cdot}64 \text{ in.} \qquad (11)$$

So that for $e=1{\cdot}35$ in. as before, $L=67{\cdot}3$ in. nearly. Hence, the higher temperature and the moisture both increasing the speed and the flatter pedal note decreasing the frequency, all concur in increasing the length of tubing required. The factors which reduce the length are therefore to be sought in (*a*) the hyperbolic shape of the tubing from the bell to the valves, (*b*) the parallel part through the valves and tuning slide, (*c*) the slight taper from the valves to the mouthpiece, (*d*) the shape of the cupped mouthpiece itself, and (*e*) the different *diminutions of the speed of sound* in each part, because the air is not open but in a small pipe. And these factors must act so as to reduce the length from 67·3 to 54·5 in.

It must not be assumed that pipes of a different proportion, like those of the euphonion say, would show the same departure as above from what might be expected as to the length from considerations of actual temperature, or that all the disturbances in this case would mutually cancel. Probably in all these forms of tubing the lengths have been settled as a matter of trial and error, tradition, etc.

(vii.) *Temperature and Pitch of Wind Instruments.*—It is evident from the foregoing that the pitch of wind instruments depends somewhat upon the temperature of the room in which they are played. But this dependence must be slight for a small instrument easily warmed by the player's breath (like the flute) and more pronounced for a large instrument, like the bombardon or the organ, where the player's breath has little or no effect. The details given in Table X. are taken from lectures delivered by D. J. Blaikley at the Royal Military School of Music.

Reeds, on the other hand, flatten with a rise of pitch. Hence, if the temperature of the air used in an organ on some occasion is much warmer or cooler than usual the reed and flue-pipes will be oppositely affected and cannot be used in combination.

Table X

Temperature and Pitch

Instrument.	Rise of Pitch due to Rise of 10° F. in Room.	
	Percentage Increase of Frequency.	Actual Increase of Number of Vibrations per Second in B♭ (479 per sec.).
Flute and Oboe .	0·31	1·50
Clarinet . . .	0·43	2·06
Cornet and trumpet . .	0·51	2·45
French horn . / Trombone .	0·60	2·88
Euphonion . .	0·66	3·16
Bombardon . .	0·73	3·50
Mean of full wind band . . .	0·54	2·60
Organ flue-pipes.	1·05	5·04

(viii.) *Free-Free Bars.*—The calculation of the frequency of lateral vibrations of uniform bars is very complicated, only the result can be quoted here. If N_1 be the frequency of the fundamental of a steel bar of length L cm. and thickness *a* cm., then it can be shown [1] that

$$N_1=538{,}400\times a\div L^2 \text{ per sec.} \quad . \quad (12)$$

Thus for a steel bar 29 in. long and $\frac{1}{2}$ in. thick we find

$$N_1=538{,}400\times \frac{2{\cdot}54}{2}\div(29\times 2{\cdot}54)^2=126 \text{ per sec.},$$

which is the value experimentally found for the prime tone of the bar in question.

(ix.) *Fixed-Free Bars.*—We meet with an approximation to this case chiefly in the prongs of tuning-forks. If of steel, we have for the frequency N_1 of the prime executed by a uniform bar of thickness *a* cm. and length L cm., the expression

$$N_1=84{,}590\times a\div L^2 \text{ per sec.} \quad . \quad (13)$$

Hence for a prong of steel 11 in. long and $\frac{1}{4}$ in. thick, we find

$$N_1=84{,}590\times \tfrac{1}{4}\div(11^2\times 2{\cdot}54)=68{\cdot}8 \text{ per sec.},$$

which is in agreement with experiment.

(x.) *Temperature-Change of Pitch in Forks.*—Since the frequency of a tuning-fork depends on its dimensions and the density and elasticity of the steel, it is scarcely likely that the value of its frequency will escape unchanged when all these factors are changed by an alteration of temperature. The variation with temperature is too small to be important in ordinary musical practice. But, as a matter of strictness, it needs notice for the sake of accuracy in copying forks or comparing one with

[1] See Rayleigh, *Sound*, i. 273.

another. If N_t is the frequency at $t°$ C. and N_0 that at 0° C. we have the relations

$$N_t = N_0(1 - 0{\cdot}000112\,t°). \quad . \quad . \quad (14)$$

The change per 1° F. would be only 5/9 of the above decimal fraction or 0·000062. This is equivalent to about one vibration in sixteen thousand per degree Fahrenheit. Thus the Kneller Hall B♭ fork, giving at 60° F. the military pitch of 479·3 per second, has at other temperatures the values shown below.

Temperature of fork	40° F.	50° F.	60° F.	70° F.	80° F.
Frequency of fork	479·9	479·6	479·3	479·0	478·7 per sec.

§ (53) Experimental Determination of Frequencies.—We shall notice first a few comparative methods that yield the difference or ratio of the frequencies and then pass on to methods that obtain absolute values of the frequency.

(i.) *Difference by Beats.*—When two musical sounds of nearly equal frequencies are about equally intense and remain so we may notice the phenomenon of *beats*, which is the alternate waxing and waning of the sound. Thus, if we start with two tuning-forks of the same pitch, bow each and place them together with the mouths of their resonance boxes side by side, we shall find a smooth flow of the sound as though there was only one fork sounding. Next, let a little soft wax be stuck on the end of the prong of one fork and some lead shot embedded in the wax. Then, on bowing the forks, an alternate loudening and softening of the sound will be noticeable. If the number of throbs or *beats per second* be counted, this will give the *difference of frequencies* of the two forks in their state at the time: for the loudest phase of the beats corresponds to the instant when the forks are impressing the ear with vibrations just in step with one another. The softest phase of the beats corresponds to the instant when the vibrations at the ear from the two forks are just out of step. To hear beats to advantage from strings and forks they should be plucked or struck and *let go.* If one continues bowing the vibrations are liable to be changed by the act of bowing and the beats will not be trustworthy, if indeed audible.

(ii.) *Detection of Mine Gases by Beats.*—The method of beats has a useful application in the detection of deleterious gases in coal mines. A portable equipment has been devised to be carried in the pit, by which two similar pipes are blown so as to emit simultaneously their notes, one pipe taking pure air from a reservoir, the other pipe being blown from the air, etc., present in the mine. Long before the air is bad enough to be dangerous its vitiation is in this way detectable by the beats between the two pipes.

(iii.) *Ratio by the Ear.*—Very often only a rough gauge of the interval between two musical sounds is required, and for this purpose the *musical ear* (where possessed) may suffice even when the sounds occur successively. The result of the test can be announced almost instantly if the interval is one in musical use. But if it is not of that simple musical character, it may take longer to decide quite what it is. The accuracy of results obtained in this manner should be quite right as to the number of semitones in a musical interval. Such decisions could be scarcely trusted, however, to discriminate between the large and small tones of just intonation, their ratio frequencies being 9/8 and 10/9 respectively. When the sounds occur simultaneously the sensitiveness of the ear is distinctly greater, and fairly good tuning is possible in this way, especially as the occurrence of beats often assists in the final judgment.

(iv.) *Forks compared by Smoke Traces.*—A convenient way of comparing frequencies of forks is that of letting each write a wavy trace on smoked glass or paper. This method has the advantage of being independent of hearing altogether. Each fork has a small style of aluminium foil fixed on it, and the forks must be so arranged that the two traces can be obtained side by side for comparisons. This may be accomplished in a variety of ways. One plan is to have both forks fixed by their shanks and some wood packing in a vice so that the four prongs are all side by side horizontally, the styles being on the underside of the prongs. Then, the forks are set in vibration, preferably by bowing, and the smoked glass moved by hand parallel to the prongs and in contact with their styles. Another plan is to have the forks mounted on a frame along which the smoked glass can slide. When the traces are obtained their wave-lengths are compared with whatever nicety is desirable for the purpose in view. Thus the ratio of the frequencies is found, since $N_1\lambda_1 = N_2\lambda_2$, where the N's are the frequencies and the λ's the corresponding wave-length of traces. Both these frequencies are slightly different from those of the forks without the styles. If the latter are sought the method of correction at end of § (53) (viii.) may be adopted.

(v.) *Lissajous' Figures.*—This is an optical method for testing the accuracy of tuning of some simple interval (unison, octave, etc.) between two forks. The arrangement is shown in *Fig.* 36.

The light from an arc lantern passes through the focussing lens L, and is reflected at a little mirror P on the fork A, then at a mirror Q on the fork B, and finally reaches the screen at O,

if both forks are at rest. The vibration of fork A will displace the spot of light parallel to XOX′, those of fork B will displace it parallel to YOY′. Hence, if both forks are vibrating together in *unison*, a special figure is described, which may be an oblique straight line or an ellipse or circle according to the phase difference of the vibrations; but nothing else is possible if the unison is exact. If, however, the unison is not exact the figure slowly melts from one of the above figures to another, and after a time, t seconds say, returns to its original form. Let the frequencies be N and N′, the latter being the less. Then in the time t one vibration has gained a complete period on the other. So

$$Nt - 1 = N't,$$

or

$$\frac{N - N'}{N} = \frac{1}{Nt},$$

whence

$$\frac{N'}{N} = 1 - \frac{1}{Nt}. \qquad (1)$$

As an example, suppose one fork is known to have the frequency 100 per second = N say, and the figures are found to go through their cycle in 2 minutes = 120 seconds. Then

$$\frac{N'}{N} = 1 - \frac{1}{12{,}000},$$

or

$$\frac{N'}{N} = \frac{11{,}999}{12{,}000}. \qquad (2)$$

So that N′ is accurate to one in twelve thousand, or N′ = 99·9917 per second. This illustrates the accuracy possible; the interval is less than one five-hundredth of a semitone.

(vi.) *Resonance Tube.*—If a tall, narrow jar or tube be taken, and the level of water in it quickly and quietly adjusted, the air column above the water may be set to respond well to a vibrating fork held over the top of the tube. Suppose the shortest length of air column that responds is L_1, then this is like a closed organ-pipe, whose length we have seen is traversed four times in the period of the vibration. Thus this distance multiplied by the frequency N of the fork gives the speed of sound in the air at the temperature ($t°$ C.) in the room at the time. Hence we may write

$$N4(L_1 + e) = v = (33{,}200 + 61t) \text{ cm./sec.}, \qquad (3)$$

where e is the end-correction for the tube and equals 0·3 of its diameter, whence N the only unknown can be found. Let the next larger air column that responds be L_2. Then one-third the length of this must correspond to the previous L_1. Thus we have

$$N4(L_2 + e) = 3v = 3(33{,}200 + 61t) \text{ cm./sec.}, \qquad (4)$$

whence, by subtracting (3) from (4), we find

$$N = \frac{v}{2(L_2 - L_1)} = \frac{33{,}200 + 61t}{2(L_2 - L_1)}. \qquad (5)$$

This is preferable to using (3) alone, if the jar or tube is long enough to permit of finding this second place of resonance.

(vii.) *Monochord Method.*—To find the frequency of a fork it is sometimes convenient to set the string or wire of a monochord so that a certain length L of it (adjusted by a movable bridge) is in unison with the fork. This adjustment should be done first by ear, and second by beats. Finally, it may be made more exact by setting the vibrating fork with its stem on the string where it crosses the bridge. It should start the string in vibration. This last is a very delicate test of the exact unison, the vibration of the string being repeatedly started and detected by the finger. Then, the tension of the string being due to a weight of M grams and the mass of the string m grams per cm., the frequency of the fork and string being each N, we have

$$N = \frac{\sqrt{Mg/m}}{2L}, \qquad (6)$$

where g is 981 cm./sec².

(viii.) *Fall Plate.*—A very useful method for the frequency N of a fork is that known as the fall plate method and illustrated in *Fig.* 37. A smoked-glass plate is allowed to fall so that a style G carried by the fork F traces the wavy mark P, Q, R on it. As seen, the plate hangs by a thread, which is burnt between the pins HH when all is ready. At

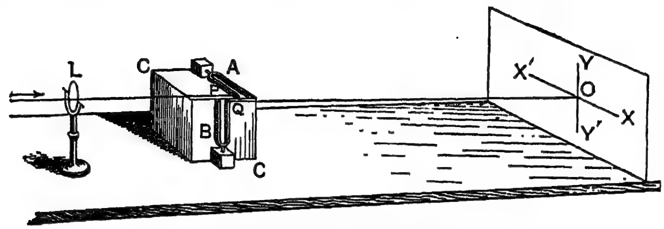


FIG. 36.—Lissajous' Figures.

the very beginning of the trace the waves are too crowded to be distinguished; but a little farther on, say at P, let them become clear. Then count an exact number of waves n, and measure the length $PQ = l_1$ totalled by them. Then, from the end Q of that set, count the same number n of waves and let them occupy the total length, $QR = l_2$ say.

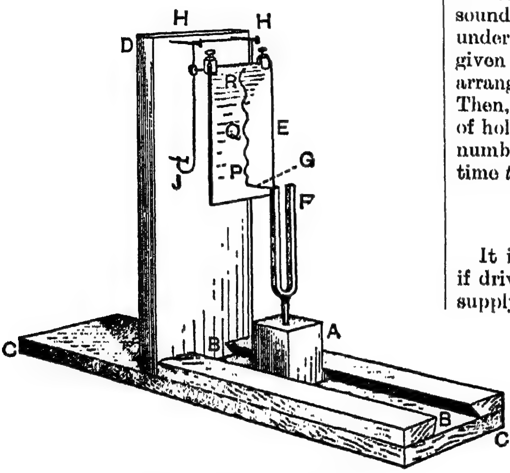


FIG. 37.—Fork and Fall Plate.

Then if u denotes the speed of fall of the plate when the style touched P, we have for the fall PQ

$$l_1 = ut + \tfrac{1}{2}gt^2, \quad . \quad . \quad . \quad (7)$$

where t is the time of n waves $= n/N$. Again, for the fall QR, the speed at Q is $u + gt$; thus we have

$$l_2 = (u + gt)t + \tfrac{1}{2}gt^2 \quad . \quad . \quad (8)$$

Hence (8) - (7) gives

$$l_2 - l_1 = gt^2 = \left(\frac{gn^2}{N^2}\right),$$

or

$$N = n\sqrt{\left(\frac{g}{l_2 - l_1}\right)}. \quad . \quad . \quad (9)$$

It should be noticed that the presence of the style on the fork will load it and flatten its pitch somewhat. This flattening could be evaluated if another fork were available exactly like that used, so that without its style they were in exact unison. Then the difference due to the style could be found by observing beats between the two forks. If it were two per second, then we should have to add two to the value found by (9) to obtain the frequency of the fork when the style was removed.

(ix.) *Siren.* If the frequency of an organ pipe is to be evaluated, some of the methods already described are not available. Indeed some other sound produced by an air blast seems preferable as a standard of comparison. For this purpose a siren may be used in which a disc with a circle of holes in it spins so that the holes are alternately opened and closed by passing a similar set of holes in a fixed plate, an air blast giving the sound. When the siren is set rotating at such a speed that its sound is in unison with that of the pipe under test, the number of revolutions in a given time must be observed by the counting arrangement on the siren and a stop-watch. Then, if the frequency sought is N, the number of holes in the circle on the disc is m, and the number of revolutions of the disc is n, in time t seconds, we have

$$N = \frac{mn}{t}. \quad . \quad . \quad . \quad (10)$$

It is not possible to keep the siren steady if driven by air alone unless some mechanical supply is available and a screw-clip or similar adjustment is used on the india-rubber pipe conducting the air to the siren. Another method is to supply the siren from a wind-chest in which the air is kept at a constant pressure by means of a loaded safety-valve. If it is not possible to keep the siren in unison, we may keep it a little too slow all the time, t, and count the total number, b, of beats thereby produced. Then, obviously, (10) changes to

$$N = \frac{mn + b}{t}. \quad . \quad . \quad . \quad (11)$$

(x.) *Tonometer.*—This arrangement consists of a set of tuning-forks extending throughout an octave or through a number of octaves. The successive forks differ in frequency by a few vibrations per second only, say, about four, all these differences being observed by beats and carefully noted, till the series building up an *exact octave* is dealt with. Then, if the bottom fork of this octave has a frequency N, the top fork has a frequency 2N. Further, let the number of beats per second between successive forks be a, b, c, d, etc., and let the sum of all these through the octave be found. Then obviously this sum is the value of N. Then the frequency of one fork being known, that of any other in the series is known by adding the beat-numbers. We can then compare any other sound with the forks of neighbouring frequencies, note the beats, and thus ascertain the frequency of the sound in question. Thus, suppose we have forks of frequencies 132, 136·1, 140·2, 143·8, and 147·9 per sec. Then, if our sound to be tested gives about 5 beats per second with the first fork, 1 beat per

second with the next one, and 3·1 beats per second with the third, and about 7 per second with the fourth, we see that the frequency sought is 137·1 per second. Such a set of forks, ranging over a number of octaves, may be taken into a belfry and so the exact pitches of the partial tones of any or all the church bells there be ascertained.

IV. Sound Detectors

§ (54) Air Resonators.—The air in a globular or conical gas or lamp shade may sometimes be observed to ring or resonate in response to some sound made by chance near it. In such cases it is often found that a musical sound a semitone higher or lower awakes very little response. Thus, the response of such an arrangement would act as a detector of the presence of a sound of particular pitch mixed up with other sounds of different pitches. To apply this principle we may use a jar, bottle, or flask and tune it to a particular pitch (as given by a tuning-fork) by pouring water into the flask to sharpen it, or partly shading its mouth by a cardboard to flatten it. In this way the first bottle picked up may be readily tuned to respond well to any ordinary fork. Then, in the case of a tuning-fork an octave lower but very vigorously bowed, this improvised resonator may be used to detect the presence of the octave of this lower fork's prime. For with large amplitudes of a big tuning-fork the air near its prongs executes a complicated motion in which the octave may be detected, although the prongs do not give that octave by simple subdivision of vibrating segments as a string does. Such a resonator may also be used to detect the octave or higher partials in the human singing voice. The use of a special form of resonator for a determination of pitch is given in § 53 (vi.).

§ (55) Sensitive Flames.—Various flames have been found to be sensitive to sounds of high or medium pitch and have been studied from several points of view. Apparently this phenomenon was discovered first by J. Leconte [1] and then independently by W. F. Barrett in 1865.[2]

(i.) *Tyndall's Vowel Flame.*—This is a high-pressure gas flame sensitive to very high-pitched sounds, and so can detect and exhibit peculiarities in certain sounds owing to the presence in them of different high partials. The gas issues from the single small hole of a Sugg's steatite pinhole burner, and the flame is about two feet high. The gas may be derived from a weighted gas bag or a steel cylinder of compressed gas with regulator. The flame must be adjusted to the sensitive state, as tall as possible without flaring. This flame responds, by falling down to a half or a quarter of its normal height, when excited by suitable sounds, such as a tap on a distant anvil, the shaking of a bunch of keys, the jingling of coins, the creaking of boots, the crumpling or tearing of paper, or the ticking of a watch. A loud *oo* leaves it unaffected, it quivers to *o*, is strongly affected by *ee*, and still more so by *ah!*

(ii.) *Rayleigh's Sensitive Flame.*—The late Lord Rayleigh devised an enclosed jet sensitive to sounds of ordinary pitches, say those through the compass of a piano. It has the advantage of working off the ordinary gas supply. A jet of coal-gas rises from a steatite pinhole burner placed in a chamber (which may be cubical of about two inches side). The gas then passes up through a vertical tube $\frac{3}{8}$-inch diameter and $6\frac{1}{4}$ inches high, mounted on the chamber. On reaching the top of the tube it burns in the open air. The front of the chamber is covered with tissue paper to receive and yield to the sound and affect the jet inside. To adjust the flame to sensitiveness, begin with the full pressure of the ordinary gas supply and slowly turn the tap down till the flame suddenly assumes a fluttering lop-sided appearance, being at one side drawn down the inside of the tube a little. Lower the pressure yet more till the flame, though still lop-sided, is steady. It is then right for use and will be found sensitive to the crumpling of paper, clapping of hands, and the piano. Its recovery after stimulus is rather slow.

(iii.) *Bunsen-burner Sensitive Flame.*—For this purpose the Bunsen burner must be carefully chosen. The following type has been found suitable: upright tube of brass 5 inches high and $\frac{3}{8}$-inch bore, only *one* side hole for air, which is *perfectly* closed by a *half*-turn of the sleeve. To obtain the sensitive state exclude the air completely and reduce the pressure of the ordinary gas supply until the flame is lop-sided but quiet. The maximum pressure consistent with these conditions seems to give the best results. The flame is about 4 inches high, that side of its base next the supply tube being detached from the lip of the upright tube and extending down into it about a third of an inch. The burner is *not*, however, *lit back*. When responding the flame falls to about $1\frac{1}{2}$ inches high and quickly recovers. It is insensitive to crumpling of paper and to the jingling of keys, but responds promptly to clapping of the hands, shuffling the feet on boards, coughing, speaking, whistling, or singing. It is possible to whistle a slow staccato passage, each note being acknowledged by a lowering of the flame and each rest by a recovery to the normal height. This flame is very useful for

[1] *Phil. Mag.*, 1858, xv. 335-339.
[2] *Ibid.*, 1867, xxxiii. 216-222 and 287-290.

detecting the nodes and antinodes where a sound is reflected from a wall. The wave-length can thus be found and the frequency of the sound calculated.

§ (56) THE RAYLEIGH DISC.—It is found that any flat obstacle in a stream of fluid tends to set itself across that stream. This was utilised by the late Lord Rayleigh in the form of a disc suspended by a fibre with its zero position at angles of 45° to the directions of flow of air in the vibrations to be detected. The quantitative treatment of such a disc was given by König in 1891. The turning couple or torque is thus shown to be expressed by

$$G = \tfrac{4}{3}\rho a^3 W^2 \sin 2\theta,$$

where G is the couple, ρ the density of the air (or other gas) moving at speed W, a is the radius of the disc, and θ is the angle between its normal and the direction of the wind W when undisturbed by the disc. A simple experiment with the disc is as follows: Just inside the mouth of the resonance box of a tuning-fork suspend by a cocoon fibre a disc of mica, thin card, or very thin mirror about an inch in diameter. Arrange that the zero position of the disc shall be at 45° with the plane of the mouth of the box. Then on bowing the fork, or singing its note in front, the disc will promptly set itself across the mouth. By light reflected from the mirror the effect may be made evident to an audience, or by use of a scale in addition the action may be used as an absolute measure of the intensity of the sound in question. In this quantitative manner the Rayleigh disc has been used in various researches in architectural acoustics and other investigations.

§ (57) HUMAN HEARING.—This is a vast subject taken in all its possible aspects; only a few of the physical ones can be noticed here. But their importance is very great, because almost all sounds are brought to the arbitrament of the human ear, hence the necessity of gaining what insight we can into the organ itself and its manner of working.

(i.) *The Ear.* A general view of the human ear, somewhat departing from exact scale for the sake of clearness, is given in *Fig.* 38.

In this we easily recognise the external ear, the ear passage ending at the drum-*skin* (often improperly spoken of as the drum). Beyond this lies the cavity which is properly called the drum. From the drum the Eustachian tube proceeds to the pharynx. This tube opens when swallowing occurs and so sets the pressures equal on each side of the drum-skin. This equality is essential to hearing, and if by plunging or diving under water, or ascending in an airplane, the equality is lost, it may be restored by purposely swallowing. The cavity of the drum is bridged across from the drum-skin to the labyrinth by a train of three little bones called respectively the hammer, the

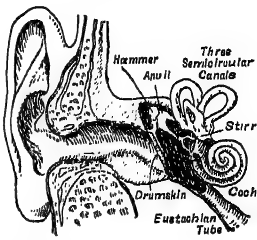


FIG. 38.—General View of Human Ear.

anvil, and the stirrup. The labyrinth comprises the vestibule, the three semicircular canals, and the cochlea. The vestibule is the middle part of the labyrinth, and contains the oval window which receives the foot of the stirrup. Forward and downward from the vestibule we have the important structure resembling a snail-shell and called the *cochlea.* It contains a spiral canal, which is seen in the cross-section of a single turn or whorl given in *Fig.* 39.

The cochlear canal (*canalis cochlearis*) is seen to be almost triangular in cross-section, and is divided from the *scala vestibuli* by the membrane of Reissner, and, by the basilar membrane, from the *scala tympani*, which ends in the round window when it reaches the

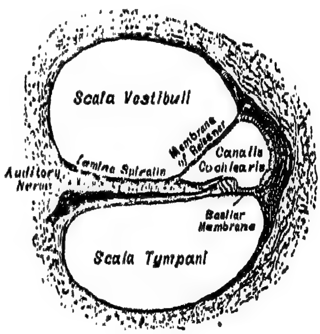


FIG. 39.—Cross-section of a Whorl of the Cochlea.

drum. These two *scalae* (or staircases) are separated by the *lamina spiralis*, through which the fibrils of the auditory nerve are distributed to the sensitive structures of the basilar membrane.

(ii.) *The Act of Hearing.*—When minute but rapid alternations of air pressure of suitable frequency and range occur at the external ear, they pass along to the drum-skin and set it in vibratory motion. This motion passes through the chain of tiny bones (hammer,

anvil, and stirrup) to the oval window, whose displacements, according to Helmholtz, never exceed a tenth of a millimetre. But the area of the drum-skin is fifteen to twenty times that of the oval window. Thus, the train of bones receives the motion of the comparatively large drum-skin moved by the *air* and changes it into the more forcible motion of the very small oval window, which has to move the *liquid* in the labyrinth. Hence the motions are handed on to this liquid, and the waves passing along the *scala vestibuli* are felt in the *canalis cochlearis* and then affect the sensitive apparatus on the basilar membrane. At the tip of the cochlea the partition ceases between the *scala vestibuli* and the *scala tympani*, so the waves proceeding along the first can return by the second. This ends at the drum with a round window which seems to play the part of a safety-valve. Much controversy has raged as to which are the structures that form the sensitive part of hearing and how they act. Some hold that these structures are some tuned vibrators which are set in sympathetic vibration by impulses of their own natural frequency. On this view, a vibration of 100 per second, say, would set in full vibration any vibrators which had the same or a near frequency (sharper or flatter), these vibrators would stimulate the nerve fibrils, and so the message would reach the brain that certain vibrators were in motion. This would constitute hearing a sound of frequency 100 per second. Any other audible frequencies would be dealt with in like manner, each setting in action those responders of frequencies near to that of the external stimulating sound. And all these could be heard simultaneously as separate sounds if not too close in frequency, or some of them too faint to be recognised amid the din of the others. Some reject this so-called resonance theory of hearing, but have little to put in its place which is intelligible or acceptable to a physicist. It is perhaps too early to feel certainty with respect to any hypothesis of audition. But it may be pointed out that one objection to the resonance theory has been recently removed. It was at ne time imagined that thousands of differently tuned responders would be needed in the ear to satisfy the requirements of that theory. It has now been shown that about *twelve to the octave* in the musical range of hearing, and, say, a *hundred* in all, may be taken as sufficient to account for all the delicacy of perception of pitch possessed by us.[1]

(iii.) *Limits of Audition.*—Separate impulses may have a frequency too slow to be heard as a continuous sound, and vibrations may be too rapid to be heard at all. Thus for each person there are limits of pitch outside which no sound is heard as such. These limits are very different for different persons, and vary with time for the same person. Some cannot hear much continuous tone for less than thirty per second, though others may hear down to fifteen per second. Some gigantic organ pipes are effective by a series of throbs shaking the floor or panes of loose glass rather than by any continuous influence they may have on the ear. The highest pitch audible is also somewhat uncertain. Thus, some can hear sounds up to or beyond 40,000 per second, while others fail at an octave lower. We might then place the extreme limits at 15 to 40,000 per second, those useful for musical purposes at about 40 to 5000 per second. Thus, while the eye only sees one octave, the ear hears about eleven octaves, of which about seven are musically available. It should be noticed that not only are the qualities unpleasant at the extreme pitches, but that the sounds are useless musically for another reason, namely, that the discrimination of pitch is obscured and lost beyond certain points both up and down. And this is just what would be the case if we heard by means of a set of vibratory responders. For obviously such a set must have its limits up and down. Further, any note *within* the range of this set would be located in pitch quite distinctly. But, beyond the range of this set, if upwards, the note would be recognised as high because only the high-pitched responders would be affected. But the pitch could not be precisely discriminated, because there would be no other responders still higher in pitch than the note heard. And to locate a pitch, on this view, it would need some responders too high to be disturbed by the sound, some responders too low to be disturbed by it, and some between that were powerfully affected. Thus a sound so high as to be beyond the range of the vibrators present would be known to be high but not clearly discriminated in pitch. Similarly, a sound lower in pitch than any of the responders present in the ear would be acknowledged to be low, but its exact pitch would elude discrimination. And this is just what we actually experience at each end of our auditory range.

It is also a matter of some interest to ascertain the total number of vibrations needed to enable an observer to decide the pitch of a sound heard. Kohlrausch investigated this by attaching to a pendulum an arc of a circle with a limited number of teeth on it. On letting go the pendulum the teeth struck a card, and on hearing the sound a monochord was set to agree as closely as possible with the pitch as it was judged to be. As the number of teeth was reduced so was the fineness of judgment as to the pitch heard. Thus, the error of judgment rose from 0·14

[1] *Phil. Mag.*, 1910, xxxviii. 164-173.

of a semitone with 16 teeth to 0·5 of a semitone with 2 teeth. Hence, even with two vibrations the judgment of pitch was only in error by a quarter of a tone.

(iv.) *Minimum Amplitude Audible.*—So far back as 1870, Toepler and Boltzmann made an estimate on this subject by experiments on an organ pipe. Their result was that with an amplitude (or displacement each side of the mean position) of 200×10^{-8} cm. a sound of frequency about 180 per second was just audible. The late Lord Rayleigh carried out experiments on two plans, one with a whistle, the other with a tuning-fork. The first, with a frequency of 2730 per second, gave an amplitude of 8×10^{-8} cm. as being just audible. The second, with a frequency of 256 per second, gave a minimum audible amplitude of $12{\cdot}7 \times 10^{-8}$ cm. Dr. P. E. Shaw in 1904 applied his electrical micrometer to test the minimum audible amplitude of the diaphragm of a telephone receiver. For an expected sound he found the minimum audible motion of the diaphragm to be 7×10^{-8} cm., and inferred that the motion of the air produced by it at the ear was only one-fifth the above. Thus the minimum audible amplitude would be $1{\cdot}4 \times 10^{-8}$ cm., or about 1 : 180,000,000 of an inch!

(v.) *Perception of Sound Direction.*—In 1907 the late Lord Rayleigh published a series of experiments undertaken to ascertain the powers of the ears in estimating the directions from which sounds came, and how these powers may be explained. By theory and experiments it was concluded that for sounds higher in pitch than about *c''* (512 per second) the discrimination of right and left is made chiefly upon the difference of intensities at the two ears, but that at low pitches, at any rate below *c* (128 per second), the recognition of phase-differences at the two ears must be appealed to. But it is to be noticed that we have very little judgment about front and back in attempting to locate the source of a sound heard. Thus, on board ship, if wishing to locate the source of a fog-signal, a combination of several observers facing different ways offers advantages. In comparing their judgments attention should be paid only to which of his sides, right or left, each listener supposed the sound source to lie.

V. Sound Reproducers and Recorders

§ (58) The Telephone. In 1876 Graham Bell patented his speaking telephone.[1] At this time his instrument acted both as a transmitter (giving to the line signals corresponding to the sounds spoken into it) and as a receiver (which reconverted those signals into sounds at the receiving end). In essential principle, Bell's telephone, as a receiver, is in use to-day; though alterations have been made in details. But, as a transmitter, the Bell telephone is superseded by some form of carbon transmitter. The Bell telephone is shown in section in *Fig.* 40, and an Edison carbon transmitter is shown in *Fig.* 41.

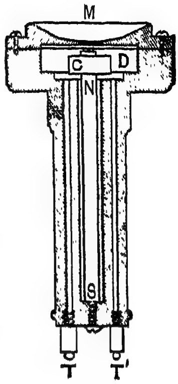


Fig. 40.
Bell's Telephone.

When using for telephony a carbon transmitter, a battery is needed to generate the current, and it is the function of the transmitter, by the variation of its own resistance, to vary the current thus produced. How this is done may be seen from the figure. The sound waves pass through the mouthpiece M and set the diaphragm D in vibration. This motion acts upon the button B and the adjoining platinum plate, which thus makes a variable contact with C, a disc of carbon (sometimes now replaced by separate carbon balls). The electrical circuit in the instrument is from the terminal T through the spring S to the platinum plate, thence through the carbon and the case of the instrument to the second terminal T'. The variations of the electrical resistance (due to the movement of the diaphragm upon which sound waves fall) are thus impressed upon the current through the line to the other station, and there act upon the telephone receiver (see *Fig.* 40). The current enters and leaves by the terminals T and T', and so passes through a coil C near the end of a magnet NS. The variations of the current attract in varying degrees the diaphragm of the receiver. It thus vibrates in correspondence with that of the transmitter, and so emits sounds which agree in essentials with those falling upon the transmitter's diaphragm.

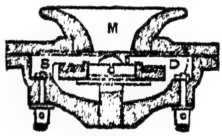


Fig. 41.
Edison Carbon Transmitter.

Portable telephone sets (called audiphones) have been arranged to relieve partially deaf persons, the carbon transmitter being on the dress and the receiver or receivers applied to one or both ears. To some patients they have proved beneficial. But at present the details of a scientific test of defective hearing seem to be not so far advanced as the corresponding ones applicable to a test of eyesight.

[1] See also "Telephony," § (17), Vol. II.; "Microphones," Vol. II.; "Microphone, The Hot Wire," Vol. II.

§ (59) Duddell's Speaking Arc.—This arrangement may be said to utilise the arc light as a special form of telephone receiver. The arc is at the same time (i.) part of the direct-current circuit which supplies its main current and (ii.) part of an oscillating circuit containing a condenser and a coil. On this oscillating circuit are impressed the alternating currents obtained by induction from the fluctuating currents of a telephone transmitter. The arc then changes its state in such wise as

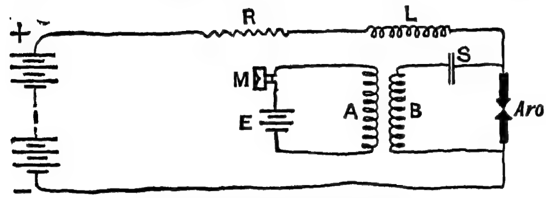


Fig. 42.—Duddell's Speaking Arc.

to emit very feeble sounds, but it can yet render speech and song and the performance on musical instruments so as to be just audible to an audience of several hundreds. The arrangement is shown diagrammatically in *Fig.* 42. It is seen that in the direct-current circuit of main supply is included an inductance L to prevent any appreciable part of the alternating current obtruding there. Also in the oscillating circuit a condenser S is included to prevent the direct current from flowing there. Only in the arc itself are both the steady and the alternating currents allowed to flow. The alternating currents are inductively produced in the coil B by the variations of current in the coil A through the waves of speech, song, etc., falling upon M the microphone (or telephone transmitter).

§ (60) Miller's Phonodeik.—Various arrangements have been adopted by different experimenters to analyse the motions of the strings and other parts of musical instruments, separately or simultaneously. Probably the first was the vibration microscope of Helmholtz. Others have involved the photographs of an illuminated portion of the vibrating string together with a uniform perpendicular motion of the sensitive film or plate. They also include the use of rocking mirrors reflecting a spot of light and so recording photographically the motions of the bridge, belly, and air of a violin, etc. These methods have been used by F. Krigar-Menzel and A. Raps, by E. H. Barton and by C. V. Raman.

In 1912 D. C. Miller published an account of his apparatus, for the demonstration and photography of sound curves, called the *phonodeik*. This is shown diagrammatically in *Fig.* 43. It collects from the air what we may call the total resulting sound from all parts of the instrument under test. This passes down the horn *h* to a diaphragm of glass about three-thousandths of an inch thick, held lightly between soft rubber rings. To the middle of the diaphragm a few silk fibres (or a platinum wire 0·0005 inch in diameter) are attached, and after passing once round a tiny pulley finish at a spring tension piece. This pulley is on a spindle carrying a mirror *m*, 1 millimetre square, the whole mass being under one five-hundredth of a gram. Light from a lamp *l* passes to the mirror focussed by a lens on to a film *f* moving in a special camera. The uniform motion of the film is perpendicular to the motion on it of the spot of light due to the vibrations to be recorded. Thus a displacement-time graph is obtained. The apparatus can be arranged also with a rotating mirror and screen instead of the camera, thus projecting the curves in the sight of an audience listening to the sounds which produce them.

Many interesting results have been obtained with the phonodeik, a few of which are reproduced here. *Figs.* 44 to 48 show records of the violin, flute, oboe, piano, and bass voice respectively. Finally, *Fig.* 49 is from a gramophone and shows Tetrazzini singing alone

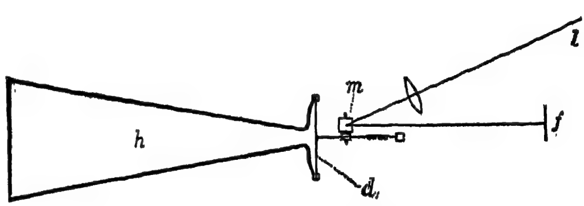


Fig. 43.—Miller's Phonodeik.

a high B♭ in *Lucia di Lammermoor*, the fluctuations in amplitude being due to a slight *tremolo*. The whole record for this one note lasts about three seconds; only about a tenth of it, just at the finish, is here shown.

§ (61) Rankine's Photophone for Transmission of Speech by Light.—In 1919 A. O. Rankine published an account of his arrangements, which may be described briefly as follows. Light from a point source is collected by a lens and an image formed on a small concave mirror which is attached to the diaphragm of a gramophone recorder. The light diverges and passes through a second similar lens, which projects it to the distant station. Two similar grids are mounted, one

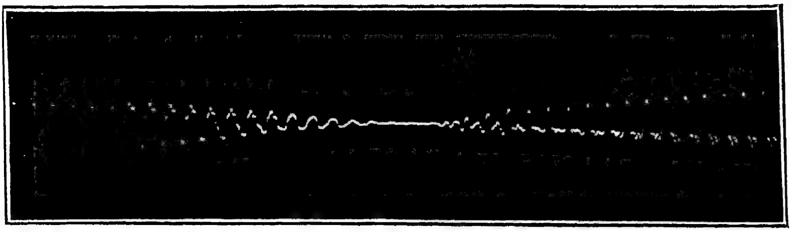


FIG. 44. - Photograph of the Tone of a Violin at the time of Reversing of the Bowing.

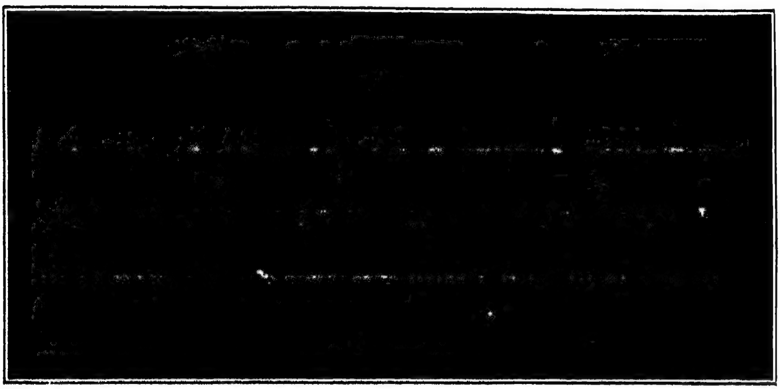


FIG. 45. —Three Photographs of the Tone of a Flute, played *p*, *mf*, and *f*.

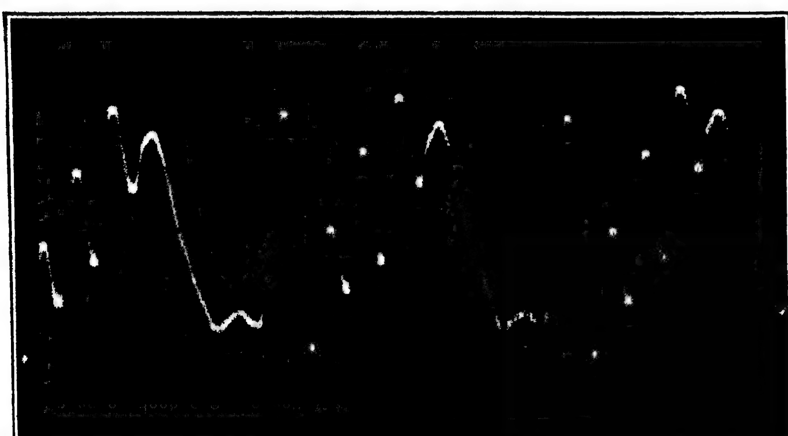


FIG. 46. Photograph of the Tone of an Oboe.

in front of each lens. An image of the first grid is superposed on the second by reflection in the concave mirror when silence reigns. But when the mirror oscillates under the vibrations of speech, the dark spaces of the image move over the openings of the second grid, thus producing fluctuations of the intensity of the beam. The light is received by a collecting lens and focussed on a selenium cell in circuit with a battery and telephone receiver.

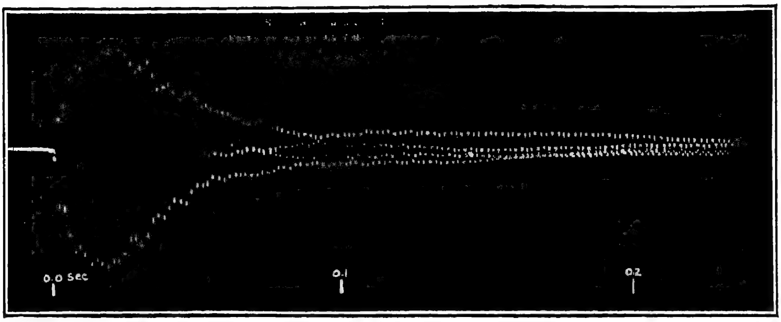


FIG. 47.—Photograph of the Tone of a Piano.

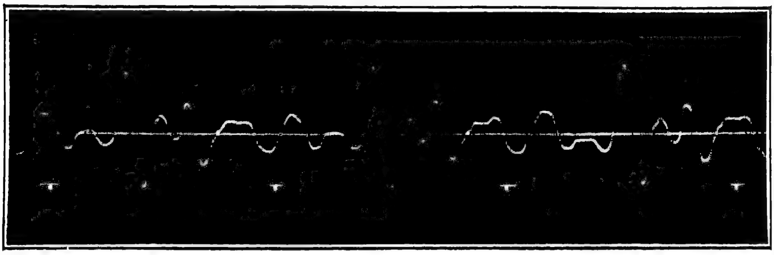


FIG. 48.—Photograph of a Bass Voice.

FIG. 49.—Sound Waves from the Soprano Voice (Tetrazzini) singing high B♭.

If the light from the above *photophone* transmitter is concentrated on a narrow slit, an image of which is produced by another lens on a cinematograph film, then the variations in intensity of the light (caused by the sounds on the photophone) are recorded as variations in the density of the film after development. For reproduction of the sounds light from an illuminated slit is focussed on the film, behind which a selenium cell is placed connected to a telephone circuit in which may be heard the sounds.[1]

E. H. B.

VI. BIBLIOGRAPHY

Books

Barton, E. H., *Text Book of Sound*, London, 1922. 687 pages. A general treatise, experimental, musical, and mathematical.

Capstick, J. W., *Sound*, Cambridge, 1913. 296 pages. General elementary treatment of the subject.

Helmholtz, H. von, *Sensations of Tone*, translated by A. J. Ellis, London, 1895. 576 pages. Very full scientific treatise of the musical aspects of sound.

[1] See *Phys. Soc. Proc.*, Aug. 15, 1919, xxxi. (v.) 242-264, and Feb. 15, 1920, xxxii. (ii.) 78-82.

Miller, D. C., *Science of Musical Sounds*, New York, 1916. 286 pages. Specially full on the graphical aspects of the subject.
Raman, C. V., *Vibrations of Bowed Strings*, i., Calcutta, 1918. Researches on the subject with photographs of vibrations.
Rayleigh, late Lord, *Theory of Sound*, London, 1896. 480+540 pages. An encyclopaedia of the mathematical aspects of sound.
Poynting and Thomson, *Sound*, London, 1904. 163 pages. Experimental treatment.
Scripture, E. W., *Study of Speech Curves*, Washington, 1906. 163 pages. Research work with numerous graphical records.
Tyndall, J., *Sound*, London, 1895. 464 pages. Experimental lectures.
Wood, A., *Physical Basis of Music*, Cambridge, 1913. 163 pages. Popular treatment.

Journals

Nature, "Mechanism of Speech," June 7, 1917, pp. 285-287; "Analysis of Musical Tones," January 13, 1921, p. 632; January 20, 1921, p. 664; March 3, 1921, p. 12.
Phil. Mag., "Monochord Vibrations," July 1905, December 1906, April 1907; "Violin Vibrations," August 1909, September 1910, May 1911, June and October 1912; "Pianoforte Vibrations," April 1910, October 1910, July 1911; "Range and Sharpness of Resonance," July 1913; "Forced Vibrations," August 1918, April 1919; "Resonance Theory of Audition," July 1919.
Phys. Soc. Proc., "Speech and Light," August 15, 1919, xxxi. (v.) 242-264, and February 15, 1920, xxxii. (ii.) 78-82.
Roy. Soc. Proc., "Impact of Pianoforte Hammer," April 1920, xcvii. 99-110.
Science Abstracts, A, "Architectural Acoustics and other Aspects of Sound," numerous articles in every volume.

ACKNOWLEDGMENTS

For kind permissions to reproduce figures acknowledgments are cordially tendered to the following:
Besson & Co., Ltd., for *Figs.* 14, 19, 21, 22, 23, 24, 25, 26, 29, 30, 31, 32, 33.
Boosey & Co., for *Figs.* 6, 7, 8, 9, 10, 11, 27.
C. G. Conn, Ltd., for *Fig.* 17.
Hawkes & Son, for *Figs.* 3A, 3B, 3C, 12, 13, 15, 16, 18, 20, 28.
Macmillan & Co., Ltd., for *Figs.* 2, 3, 5, 34, 35, 36, 37, 38, 39, 40.
D. C. Miller and The Macmillan Co., for *Figs.* 41, 42, 43, 44, 45, 46, 47.

SOUND, ABSORPTION COEFFICIENTS OF, for various materials, tabulated. See "Sound," § (22) (vi.), Table VI.

SOUND, LONG-RANGE TRANSIT OF. See "Sound," § (23).

Zones of silence. See *ibid.* § (23) (viii.).

SOUND, SOURCE OF. See "Sound," § (8).

SOUND, SPEED OF, affected by temperature but not usually by pressure. See "Sound," § (10).

In air; determination by Hebb, using parabolic reflectors and telephones. See *ibid.* § (20).

In air, experimental determinations. See *ibid.* § (13).

In air, increase with intensity. See *ibid.* § (14).

In iron, experimental determination. See *ibid.* § (17).

In pipes; Blaikley's determination with smooth pipes. See *ibid.* § (19).

In pipes; Blaikley's values for dry air, at 0° C., in brass tubes, tabulated. See *ibid.* § (19), Table IV.

In pipes, Regnault's determination with rough pipes. See *ibid.* § (18).

In various media, tabulated. See *ibid.* § (21), Table V.

In water, experimental determination. See *ibid.* § (15).

SOUND, THEORETICAL SPEED OF, in liquids. See "Sound," § (11).

In solids. See *ibid.* § (12).

SOUND PRODUCER, CALCULATION OF FREQUENCY OF. See "Sound," § (52).

SOUND PRODUCERS. See "Sound," III., § (24) *et seq.*

SOUND RANGING

§ (1) GENERAL.—"Sound Ranging" consists in the location of a source of sound by means of measurements made on the sound wave which spreads from the source. *Fig.* 1 illustrates the manner in which this may be done. A sound originates from the point S in the figure, and is received at three stations A, B, and C. At each station the time is registered at which the sound wave arrives, and so the time intervals between its reaching A and B, B and C, can be measured. A geometrical construction such as that shown in the figure then gives the position of the source with reference to the surveyed stations A, B, and C. If the sound arrives at B t_1 seconds after reaching A, and at C t_2 seconds after arriving at A, circles are drawn with B and C as centres and with radii equal to Vt_1, Vt_2, where V is the velocity of sound in air. The centre of a circle which passes through A and touches the other two circles will coincide with the point S. This point S may alternatively be regarded as the intersection of two hyperbolas. If the time interval for the stations A, B is t_1 seconds, all possible positions for S lie on a hyperbola, with A and B as foci, and which is determined by the relation $SB - SA = Vt_1$. A similar hyperbola can be drawn for B and C, and the two curves intersect in S.

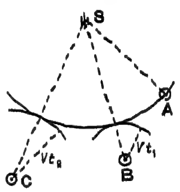


FIG. 1.

It will be clear that the method is not applicable to a source of continuous sound; it can only be used when the sound has a sharply defined commencement.

Towards the end of 1914, when the fighting in France and elsewhere developed into "trench warfare," it became clear that Sound

Ranging might provide a valuable means of locating hostile artillery. The batteries on both sides took up fixed positions, in which the guns were protected from shell fire by emplacements, and were carefully camouflaged to screen them from aerial observation. Their location became increasingly difficult. At the same time the stability of the front line made it possible to develop more complex methods of determining their positions. For instance, one method which soon proved of great value consisted in the installation of observation posts at suitable points behind the front line, from which bearings were obtained of the flashes and smoke puffs when hostile guns fired. These positions were accurately surveyed, and the locations determined by the co-operation of several posts. Experiments were also commenced in the French and German armies, and later in the British army, to test the practicability of Sound Ranging, and their successful result led to the establishment of Sound Ranging as a valuable means of location.

§ (2) Sound Ranging in Practice. (i.) *Difficulties of Location.*—Although the principle of Sound Ranging is simple, in practice many difficulties had to be surmounted. The batteries to be located were situated from 5000 to 15,000 metres from the points where it was possible to put the recording stations. In order that the locations may be of value they must be accurate to within 100 or 200 metres. A consideration of the displacement of the plotted position which is caused by errors in timing the sound shows that these timing errors should not exceed one-hundredth of a second. For this reason it is impossible to use human agency to record the sound wave, except under the most favourable conditions. A man recording the commencement of a gun report, by pressing a key or starting a watch, will make errors of at least one-tenth of a second. Some success was attained by systems of sound ranging in which stop watches, or registering mechanisms actuated by the pressing of keys, were employed; and in the German army these methods were continued until a later date than in the British and French armies. Finally, however, they were replaced by an automatic registration of the sound by means of microphones or similar devices connected by cable to a central station.

The velocity of the sound wave is affected by the wind and is dependent on the temperature, so that allowances must be made for these factors. Both wind velocity and temperature vary with the height from the ground, and measurements of them made at ground level will not serve as a basis for calculation. This question will be dealt with more fully below, for the calculation of the true wind and temperature corrections is a complex and difficult problem.

(ii.) *Effect of Wind.*—Until experiments were made, it was not generally realised to what an extent the audibility of a sound coming from a point several miles away depends on the direction of the wind. The wind velocity almost invariably increases with the height from the ground. A sound wave proceeding in a direction opposed to that of the wind is in consequence refracted upwards, its velocity along the ground being greater than that at some distance above it. Its energy is dissipated upwards, and it will not affect a microphone or observer on the ground level. It was found impossible to make locations when the wind was blowing from our side of the line towards the enemy; the reports of the guns never reached the microphones, or were drowned by other noises on our side of the line. On the other hand, when the wind was blowing from the batteries towards the recording stations, the reports were heard clearly and had a sharply marked commencement.

The wind, besides displacing the apparent centre from which the sound wave is spreading, also causes irregularities due to local variation in velocity. The sound wave is slightly distorted and cannot be represented on the plotting board by the arc of a circle. No perfection in the accuracy with which timing is made will obviate this, and the errors thereby caused can only be reduced to reasonable dimensions by having the recording stations a long distance apart. The system of recording stations, generally five or six in number, constitutes what is termed the sound-ranging "base," and it is necessary to have a base from 5000 to 8000 metres in length in order to record with accuracy guns at normal ranges.

§ (3) Sound-ranging Stations. (i.) *The Base.*—*Fig.* 2 shows diagrammatically the general arrangement of the microphones and central station which was employed on the British front. The receiving stations were six in number. They were arranged regularly along a base three or four miles in extent, at positions which were surveyed with extreme accuracy. The relative positions of the microphones must be known to a metre, and their absolute positions with relation to the trigonometrical framework to about twenty metres, if the locations are to be accurate. The base was situated about 3000 metres from the front line. Each microphone was connected by cable to a central station behind the base, and at this station a recording apparatus registered the variations in electrical current caused by the impinging of the sound on the microphones. It consisted of an Einthoven galvanometer, in which six

fine wires were mounted between the poles of the electromagnet, each wire being connected to one of the microphones. The shadows of these six wires were thrown, by a suitable arrangement of lamp and lenses, across a narrow slit in the side of a camera. A strip of cinematograph film passed continuously behind the slit while a record was being taken, and, if the Einthoven wires were stationary, the film when developed showed six continuous straight lines where the shadows of the wires had protected it from exposure to the light. If a wire quivered in response to a signal from a receiving station, the line on the film had a corresponding "break." The illumination was cut off one hundred times a second by a toothed wheel rotating in front of the lamp, the angular velocity of the wheel being regulated by a tuning-fork. In this way a series of timing lines was ruled across the film. By counting the number of lines between two breaks, the time interval between the arrival of the sound at the two stations could be read. *Fig.* 3 shows a typical record, that of the report of a howitzer which has reached all six microphones in turn. The film has run from right to left, so that the left-hand edge of each break marks the arrival of the sound. The time interval can be estimated on the film to ·002 of a second.

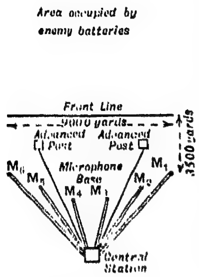


FIG. 2.

(ii.) *Observation Posts.*—In front of the sound-ranging base two forward observation posts were stationed. These were so placed that enemy guns were heard at one post or the other at least three seconds before the sound arrived at the nearest microphone. The observer in each post pressed a key when he heard a gun fire, and the key operated electrically a switch at the central station which set in motion the recording apparatus. The observer kept the key depressed until he judged that the sound of the gun and of the shell-burst had reached all the microphones. He then stopped the recording apparatus and telephoned to the central station such information about the sound as would aid the computing staff to make the location. The film was developed and fixed, the time intervals read, and the position of the gun plotted on a chart of the surrounding district. An automatic means of developing and fixing the film was finally adopted by means of which these two processes could be completed in fifteen seconds.

(iii.) *The Microphone.*—The microphone was of a special type.[1] The report of a gun differs from other sounds in that a great part of the energy is represented by waves of very low frequency. In the case of a 15-in. gun, for instance, the main oscillations of pressure occur at the rate of four or five per second. The microphone was made responsive to waves of low frequency, and was practically unaffected by the rapid oscillations of pressure caused by other noises than reports or shell-bursts. A gun-report almost imperceptible to the ear was recorded by it, whereas rifle fire, the noise of traffic, and other disturbances, produced no effect. This tuning was done by placing it in a small aperture opening into a

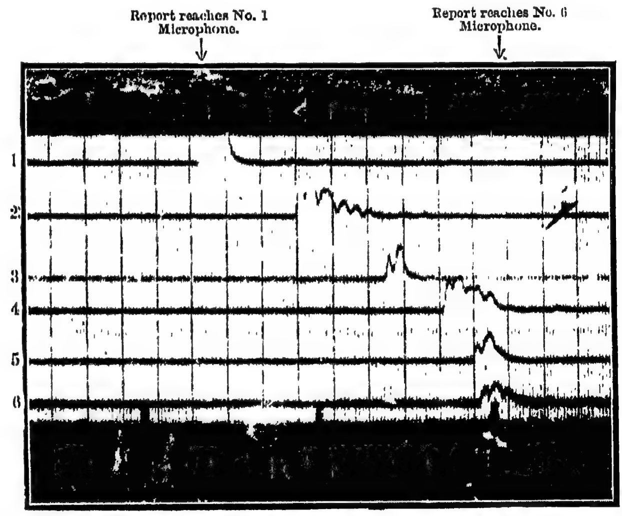


FIG. 3.—The figure is an enlarged print of the record of a 15-cm. howitzer, the report of which has reached No. 1 microphone first and No. 6 microphone last. The film has been moving from right to left while the record was being taken. The time intervals are marked by vertical lines, one hundred to the second, every tenth line being heavier so as to facilitate counting. The horizontal lines represent the shadows of the Einthoven strings, which lie across the slit behind which the film is exposed, and the movements of which are shown on the record.

[1] See "Microphone, The Hot Wire," Vol. II.

canister of large volume, on the principle of the Helmholtz resonator, constituting a highly damped system whose natural period was about one-twelfth of a second. It was further necessary to devise a type of microphone which recorded the amplitude and wave-length of the sound, and therefore gave characteristic records. Many other sounds interfere with those from the gun being located, and this made it necessary to be able to recognise on the film the breaks which belonged to the same series.

§ (4) The "Onde de Choc."—One of the most formidable difficulties in measuring the time of the arrival of the report was found to be the "Onde de choc," or "Shell-wave," which precedes the true report when the observer is within a certain area in front of a gun which is firing. The shell from a gun travels initially with a velocity greater than that of sound, and in so doing it creates in the air a conical bow-wave such as is evident in the well-known photographs of a rifle-bullet in flight. This bow-wave is heard as a sharp crack, the intensity of which, as judged by the ear, is generally very much greater than that of the report which succeeds it. The shell-wave affects the ordinary type of contact microphone to a far greater extent than does the gun-report, and generally masks the latter in consequence. The special type of microphone employed, invented by Lieut. Tucker and named after him, distinguishes between the two sounds, and the adoption of this microphone contributed very greatly to the success of Sound Ranging in the British army.

§ (5) Location of Shell-bursts.—When a record of a battery in action was obtained, it was possible to locate the bursts of the shells as well as the guns which were firing. The distance between gun and burst told the range at which the battery was firing, and a simple calculation from the time interval on the film between report and burst gave the time of flight of the shell. By comparing the range and time of flight with those for various types of guns as given in the range-tables, a guide to the calibre of the piece could be obtained. The shell-wave was also of assistance in the determination of the type of gun. No shell-wave is created by the low-velocity shell from a howitzer, so that howitzers and guns could at once be distinguished. The heavier the piece, and the greater the initial velocity of the shell, the longer is the time interval between the shell-wave and the gun-report as heard by an observer in the line of fire. This interval can be calculated from the range-tables for each gun, and when the interval read off from the film is corrected for the elevation and line of fire, it is a guide in the determination of type. It was also found that the wave-length of the sound increased with the calibre, and could be used as an indication of it. Finally, since the position of the bursts had been recorded, it was possible in many cases to tell the calibre from the fragments of shell and the fuses in the shell-craters.

This information was of great value. The number and character of the batteries on any particular section of front was an indication of the offensive or defensive intentions of the enemy in that area. Further, in cases where large numbers of batteries were placed in a limited area, locations, owing to errors, might overlap to a considerable extent. If the calibre were determined in each case, this mass of locations could be resolved into groups which were assignable each to one battery, and the probable position of that battery fixed with greater certainty.

§ (6) Wind and Temperature Corrections.—It has been stated that wind and temperature corrections play an important part in accuracy of location. If rays are drawn from the source of the sound, these rays being everywhere normal to the wave, like those used to illustrate the propagation of light, then the sound will travel from source to microphone along that path for which the time taken is the shortest possible. When the wind is blowing from the gun towards the microphone, the highest point on the path may be one or two thousand feet above the surface of the ground, since at these heights the greater wind velocity more than compensates for the longer distance over which the sound has to travel. The temperature may be very different at these heights from that near the ground. The problem of calculating the corrections is so complicated that it was found necessary to determine them empirically. At certain points, at some distance behind the front, microphones were arranged on the arc of a circle of about 7000 metres radius. Every hour a charge was exploded at the centre of the circle and the average velocity of the sound travelling to each microphone was determined directly. From these data values for the "effective wind and temperature" could be calculated which, when applied as corrections to the results obtained by sections at the front, brought their locations on to the true gun positions. The effective wind and temperature found in this way were nearly the same for all the stations on the front, except when some rapid change in the weather conditions was taking place. By comparison of these results with the meteorological data measured in the ordinary way, empirical relationships between them were formulated which could be used when the special installa-

tions to measure wind and temperature were not in action.

§ (7) METHOD OF USE.—Sound Ranging could only be employed when the front remained stationary for several days at the least. In installing a section, a base must be planned and carefully surveyed, positions made for the microphones, central station, and advanced posts; some forty or fifty miles of cable laid in an exposed situation, the apparatus set up, and the chart constructed on which the results are plotted. The time necessary to carry this out was reduced by training and experience to two or three days, and could no doubt be still further reduced by improved methods. If a method of wireless transmission of the signals could be devised, it would obviate the labour and delays caused by the constant cutting of the cable by traffic and shell-fire. Though the use of Sound Ranging has hitherto been restricted to conditions of stationary warfare, it may develop into a means of location available under all conditions.

W. L. B.

SOUNDING MACHINES. See "Navigation and Navigational Instruments," § (17) (iii.).

SPACE-LATTICES defined by Frankenheim and Bravais, fourteen in number, fundamental to the 230 modes of arranging points in a homogeneous structure such as is possible to crystals. See "Crystallography," § (11).

SPARK-GAP: the maximum sparking distance between point or sphere electrodes, the degree of separation being a measure of the applied voltage. See "Radiology," § (20).

SPECIFIC HEAT OF GLASS. See "Glass," § (28).

SPECIFIC INDUCTIVE CAPACITY OF GLASS. See "Glass," § (30).

SPECIFIC RESISTANCE OF GLASS. See "Glass," § (30).

SPECTRA:

Arc and spark, Fowler's work on. See "Spectroscopy, Modern," § (11).

Band: a type of spectrum, usually associated with the spectra of compounds or molecules, which are composed of bands; that is, of groups of lines which converge to definite heads, the head of a band being frequently the strongest line in the band. The lines in a series of bands may be represented by the formula

$$n = Am^2 + Bn^2 + C,$$

where A, B, C are constants and m and n integers. See *ibid.* § (12).

Classification of luminous, into two classes: (1) continuous spectra, due to the radiation of heated solids; (2) discontinuous spectra, in general peculiar to luminous gases. See *ibid.* § (2).

Dispersion and resolution in. See "Diffraction Gratings, Theory of," § (6).

Overlapping. See *ibid.* § (7).

SPECTROMETER. See "Spectroscopes and Refractometers."

Adjustments of. See "Spectroscopes and Refractometers," § (8).

Hilger's Constant deviation type used as monochromatic illuminator. See "Immersion Refractometry," § (7).

SPECTROPHOTOMETRY

I. INTRODUCTION

ALL sources of illumination, such as the sun, the electric arc and spark, the incandescent electric lamp, the gas and kerosene flames, the fire-fly, and countless others, emit what is known as radiant energy. When this impinges upon any obstacle and is in part absorbed, there is, in general, a rise of temperature of the absorbing substance. Other manifestations of the reception of radiant energy are also well known. Some kinds, for instance, when reaching the retina of the eye give rise to the sensation of colour, some cause blackening of a photographic plate (when it is exposed and developed), some give rise to an electric current under certain conditions, and so on. The time-rate of transfer of radiant energy is called the radiant power, and this may, of course, be expressed in the same units as other power. It has become of great importance in many branches of science and industry to be able to analyse qualitatively and quantitatively the radiant power which any source emits, or that which a given material may transmit or reflect.

When this radiant power is examined by means of a prism and the resulting spectrum, it is found that only a part is visible to the eye; that beyond the red end of the spectrum is a large invisible region, called the *infra-red*, which may be detected by means of the rise in temperature of various kinds of sensitive radiometers; and that in the other direction beyond the violet is a second invisible region, called the *ultra-violet*, which may, if sufficiently intense, be detected by these same radiometers, but much more easily by the photographic plate or photoelectric cell.

The usual types of radiometer, which depend for their action primarily upon the heating effect of the incident radiant power, are non-selective as regards wave-length or frequency, and can be used to measure this power in absolute quantities. The eye is different from the radiometer, however, in being sensitive to only a part of the radiant power from any source, that part by definition being the visible region of the spectrum. It is also different in that it cannot give the absolute

value of this radiant power. When, however, two beams of light of the same quality and brightness are brought into the proper juxtaposition, the eye can detect very small variations of brightness in either beam provided the other remains constant. Therefore, if the proper means are available for varying the brightness of either beam, the two may very accurately be brought to the same brightness. This fact is the fundamental basis of photometry.

Spectrophotometry, strictly speaking, is the *visual* measurement of *relative* radiant power as a function of wave-length or frequency. The term, however, is usually not limited to visual methods; it is often made to include also certain radiometric, photoelectric, and photographic methods of measurement used not only in the visible region but in the ultra-violet or infra-red. The term is so used herein. The kinds of work are many and varied, but may be classed, in general, under three heads:

(i.) The measurement of the *relative spectral distribution of radiant power* of sources of light. If the absolute or relative spectral distribution is known for any one source, used as a standard, the others may easily be computed.

(ii.) The measurement of the *spectral transmissive properties* of transparent materials, such as eye-protective or signal glasses, dyes, oils, and all kinds of organic and inorganic solutions—in fact, solid, liquid, and gaseous matter of great variety.

(iii.) The measurement of the *diffuse spectral reflection* of opaque objects, such as paints, cloth, paper, and so on. Magnesium carbonate or oxide is usually taken as a standard white for comparison purposes.

Spectrophotometry is generally admitted to be the *fundamental basis* of colour specification. "A spectrophotometric table, derived from at least twenty-five points (for a continuous spectrum), gives the only unique description of a colour, and it appears probable to the writer that the requirements of precision technical colour measurement are most likely to be met by the development of simple and rapid means of plotting and recording accurate spectrum plots of reflection or transmission characteristics." [1] The large and growing demand made by industrial and commercial as well as scientific interests for the standardisation and specification of colours may be partly realised by referring to a paper on "The Work of the (U.S.) National Bureau of Standards on the Establishment of Colour Standards and Methods of Colour Specification." [2] The relation between spectrophotometry and colourimetry will be brought out more fully in Part II.

[1] Ives, *Jour. Franklin Inst.*, 1915, clxxx. 700.
[2] Priest, *Trans. I.E.S.*, 1918, xiii. 38.

It has not been attempted to give a history of the science of spectrophotometry nor a complete bibliography. Such will be found, up to the year 1905, in Kayser's *Handbuch der Spectroscopie*, vol. iii. A fairly detailed description of visual instruments in use up to 1909 is also given in Krüss' *Kolorimetrie und Quantitative Spektralanalyse.* The references given herein, however, cover the field in a general way; and a study of the papers cited, together with the references which they in turn suggest, will lead one directly into a comprehensive study of the whole subject.

Most of the different types of visual instrument are mentioned herein. Since certain fundamental principles underlie the use of all of them, one of these instruments has been selected for a detailed description of method of use. The instrument and method so selected have been proved accurate and reliable by direct and extensive comparison of results with those obtained by other methods, not visual. The photoelectric and photographic methods so tested are likewise, by this intercomparison, proved accurate and reliable. They are both null methods and eliminate many possible errors, uncertainties, and difficulties. Because of their importance in confirming and extending results obtained visually, these are also described in detail. A fourth method available in this intercomparison of methods has been the thermoelectric method. This is briefly described, but the large and important subject of spectroradiometry in the infra-red is otherwise untouched. This fourfold comparison of methods has made it possible to found a statement of the accuracy at present obtainable by spectrophotometric methods upon a basis of fact rather than of opinion; but there is no reason to believe that the accuracy obtainable may not be still further increased by further work along this line.

II. Relation of Spectrophotometry to Colourimetry

§ (1) The Need for a Recognised Nomenclature and System of Standards.—For an intelligent discussion and understanding of any science there should be available an exact and well-known nomenclature and system of standards. Otherwise one is under the necessity of giving detailed definitions of all terms used if he wishes to be sure that the reader will understand by the terms what he himself has in mind. This has been largely the case in papers on spectrophotometry, especially those which have reference to the measurement of the transmissive properties of materials, which part of the science has probably had a wider field of application than the others.

The sciences of spectrophotometry and colourimetry are inseparably connected; and a majority of the terms and expressions used in one are also necessary in the other. An effort to eliminate the present confusion in the nomenclature of the science of colourimetry has been initiated by the Optical Society of America and a preliminary report made.[1] The following paragraph from the introduction to this Report is very pertinent in this connection:

In the field of chromatics and colourimetry, experiment and knowledge have outrun established and recognised language. There are many names for the same thing and many different things designated by the same name. Progress requires the co-operation of individuals and the co-ordination and discussion of results; and co-operation and discussion are primarily conditioned upon unambiguous language. While numerous pleas for standardisation and some detached recommendations for particular definitions and terms have been made in the past, no attempt at all commensurate with the magnitude and importance of the subject has been made to formulate and present a consistent organised system of nomenclature as extensive as is found necessary in practice. Probably the most urgent need of colourimetry is the establishment of a recognised nomenclature and system of standards. To one not thoroughly familiar with the subject, the insistence upon this, and particularly the elaborate and extensive system with fine distinctions of meaning to be set forth in this report, may appear pedantic and academic. Not so to him whose daily work is in this field, for in his preparation of reports and ordinary conversation with his colleagues and assistants, he is continually inconvenienced and annoyed by the circumlocution and misunderstanding occasioned by lack of suitable terms and symbols to express his ideas and findings cogently and without ambiguity. He finds that the orderly and efficient conduct of his work compels him to assume the tedious and unpleasant task of formulating a nomenclature. This necessity is the cause of the present report.

Certain parts of this Nomenclature Report are of special value for the science of spectrophotometry, and are therefore presented here. All the quotations in Part II. of the present article are taken from that Report.

§ (2) General Terminology.—"Colour is the *sensation* due to stimulation of the optic nerve."

"Colour may be evoked by various stimuli acting on the optic nerve, for example, mechanical and electrical stimuli; but the only one of these with which we shall be much concerned is *light* (see definition below), which may be called the proper or ordinary stimulus of colour.

"In adhering to the strictly subjective definition of colour we are not estopped from the use of such convenient expressions as 'the colour of light,' or the 'colour of a material,' for we proceed consistently to define:

"The colour of light as the sensation which that light evokes; and

"The colour of a material as the sensation evoked by the light transmitted or reflected by that material under specified conditions of illumination."

"Chromatics is the science of colour."

"Colourimetry is the science and practice of determining and specifying colours by reference to the stimuli and conditions which evoke them."

"Light is *radiant power* multiplied by the visibility of the radiant power in question."

"Radiation is the *process* by which energy is propagated through space. . . ."

"Radiometry is the measurement of radiant energy or radiant power." Spectrophotometry is thus seen to be one branch of radiometry.

"The Spectrum is a graphic arrangement or setting in order of radiant energy with respect to wave-length or frequency."

The *wave-length units* which are convenient for use in spectrophotometry are the *micron*, μ (one-millionth of the metre), and the *millimicron*, $m\mu$ (one-thousandth of the micron). The latter unit is used throughout this article.

"Brilliance is that attribute of colour without which colour cannot exist, the attribute in respect of which all colours may be classified as equivalent to one or another of a series of *grays* of which *black* and *white* are the terminal members."

"Hue is that attribute of colour in respect of which it differs from *gray*, the attribute in respect of which colours may be classified as *red*, *orange*, *yellow*, *green*, *blue*, *violet*, and *purple*, or their intermediates."

"Saturation is the distinctness or vividness of *hue*."

"Quality is the property of colour due to its *hue* and *saturation*."

§ (3) Methods of Specifying Colour.—"A colour may be specified by reference to stimuli in the following ways:

"(i.) By specification of the identical radiant power actually evoking the colour, both in regard to

(*a*) Its total amount,

(*b*) Its spectral distribution.

"The crude specification of spectral distribution referring roughly to spectral *regions* rather than points is sometimes used for special purposes,[2] but its limitations must be kept in mind.

"The *partial* specification of spectral distribution may be used as a colour specification in particular

[1] Optical Society of America, Committee on Nomenclature and Standards, Sub-committee on Colourimetry, L. G. Priest, Chairman, *Report*, 1919 (Preliminary Draft), U.S. Bureau of Standards Library, Washington, D.C.

[2] Ives, *Tint Photometer*, Hess-Ives Co., Philadelphia; Block, "Die Kennzeichnung der Farbe des Lichtes," *Electrot. Zs.*, 1913, xlvi. 1306.

cases,[1] but as in the case above, its limitations must be carefully observed in order to avoid erroneous conclusions."

"(ii.) By specification of a stimulus empirically found to evoke the same colour in juxtaposition.

"There are, in practice, several important special cases of this. They are:

"(*a*) The stimulus composed of heterogeneous light of the spectral distribution which alone evokes the colour gray plus homogeneous light—the wave-length of the latter, the total light, and the ratio of heterogeneous to homogeneous light all being adjusted by trial and error until the sensation evoked is exactly like the colour to be specified.[2] This method we call homo-hetero-analysis.

"(*b*) The stimulus composed of lights of three (or more) different frequencies (or frequency ranges), the relative and absolute values of which are adjusted by trial and error until the sensation evoked is exactly like the colour to be specified.[3]

"(*c*) The spectral distribution of energy of the stimulus modified by rotatory dispersion, and its total radiant power adjusted in any way by trial and error to evoke a sensation exactly like the colour to be specified.[4]

"(*d*) The spectral distribution of energy of the stimulus and its total radiant power modified by the selective transmission of glasses or solutions[5] chosen by trial and error to evoke a colour exactly like the colour to be specified.

"(*e*) The stimulus being radiant power from a complete radiator ('black body'), the temperature of which is varied until the quality of the colour evoked is exactly like that of the colour to be specified, the spectral distribution of energy being a known function of temperature."

§ (4) The Fundamental Physical Basis of Colourimetry.—"The method indicated under (i.) of § (3) is, of course, the primary basis of colourimetry." In this connection, note the statement by Ives[6] already referred to.

"Radiometry, *in its most general sense*, is the key to all fundamental work in colourimetry. Some empirical testing and comparing may be done without *direct* reference to it; and its bearing on the subject may not be obvious to the casual observer of routine testing; but the essence of all fundamental standardisation is radiometry. Note the following:

"(*a*) The fundamental reference standards of colourimetry are light sources of specified *energy distribution*.

"(*b*) The most important specification of the colour of a transparent plate is its *spectral transmission*.

"(*c*) The most important specification of the colour of an opaque object is its *spectral reflection*.

"If two transparent objects have identical spectral transmissions, or two opaque objects identical spectral reflections, they will evoke colours exactly alike *in light of any spectral energy distribution*. Moreover, this is the only form of colour specification for materials which will *ensure* this result."

§ (5) Technological Applications of Spectrophotometry.—The following examples will illustrate to what uses the science of spectrophotometry has been or may be put as regards its technological applications:

Many vegetable oils, such as cotton-seed oil, for example, have for many years been graded from the standpoint of their colour, approximate and empirical methods being used. The present status of this work, and the growing demand that the standardisation be put upon a spectrophotometric basis, may be noted by referring to recent numbers of the *Cotton Oil Press*.[7]

Accurate spectrophotometric data are of value in the analysis of dyes,[8] or would be if such were available to any great extent. The concentration of all kinds of solutions might also be determined from known transmissive data.

The great importance of securing adequate protection from injurious radiant energy is now well recognised, and a large number of glasses are on the market advertised to protect the eyes from the ultra-violet and infra-red and from excessive brightness in the visible region. Since both the amount and the wave-length of the transmitted energy determine the value of these glasses for protective purposes, it is obvious that their transmission curves should be accurately known. The results of such a study on a large number of glasses in the ultra-violet, visible, and infra-red have been published.[9] The last of these papers gives also the spectral transmission of

[1] Priest and Peters, *Bur. Stand. Tech. Pap.* No. 92.

[2] Bezold, *Theory of Colour*, Eng. trans. by Koehler, Am. Ed., 1876, p. 100; Rood, *Modern Chromatics*, 1879, chaps. iii. and xiv. p. 213; Abney, *Colour Measurement and Mixture*, 1891, chaps. iv. and xiii.; Nutting, *Bur. Stand. Bull.*, 1913, ix. 1; Jones, *Trans. I.E.S.*, 1914, ix. 687; Priest, *Jour. Wash. Acad.*, Feb. 4, 1916.

[3] Maxwell, "Colour Box," J. Clerk-Maxwell, *Sci. Pap.* i. 420; F. E. Ives, *Jour. Frank. Inst.*, 1907, clxiv. 47 and 421; Howland, *Science Conspectus Soc. of Arts, Mass. Inst. of Tech.*, 1916, vi. 29; *Carpets, Wall Papers, and Curtains*, Feb. 2, 1918.

[4] Konig, *Ann. der Phy.*, 1882, xvii. 990; Meisberg, *Zs. für anal. Chem.*, 1904, xliii. 137; Arons, *Ann. der Phys.* (4), 1910, xxxiii. 799, and *ibid.* (4), 1912, xxxix. 545; Priest, *Phys. Rev.* (2), 1917, ix. 341, and *ibid.* x. 208.

[5] Lovibond tintometer; "Arny's Solutions," H. V. Arny, *Jour. Ind. and Eng. Chem.*, Oct. 1919, xi. 950.

[6] Ives, *Journ. Frank. Inst.*, 1915, clxxx. 700.

[7] Chemists' Section, conducted by the American Oil Chemists' Society, *e.g.* Jan. 1921, pp. 42-44.

[8] Mathewson, *Assoc. Off. Agr. Chem.* vol. ii. No. 2, p. 164.

[9] *Bur. Stand. Tech. Pap.* No. 93, 1917, 3rd. ed., 1919; No. 119, 1919; No. 148, 1920.

coloured railway signal glasses and of glasses used in ultra-violet signalling.

Selective ray filters, both of glass and of dyed gelatine films, are now widely used to increase visibility (*i.e.* the ability to see clearly).[1] This is true not only visually, but has become of great importance in aeroplane photography. The increase in visibility is brought about by absorption of the shorter wave-lengths, and the spectral transmission of the glasses and dyes used is necessary to make any intelligent study of the subject. Other filters, especially the dichromatic red-blue kind, are of value in the detection of chromatic camouflage.

III. Visual Instruments

§ (6) General Characteristics.—The *visual spectrophotometer*, as the name would imply, is really composed of two separate instruments, the spectrometer and the photometer; and various problems in the construction and use of these instruments are, therefore, present in their combination. Those which have special reference to spectrophotometry as a single and separate science will be discussed at the proper places.

(i.) *The Photometric Field.*—In practically all visual spectrophotometers two beams of light (beams 1 and 2) of approximately the same wave-length are finally brought to form the two parts of what is called a photometric or comparison field, which may be viewed by the eye, and the brightness of one or both parts of which may be varied at will in a known manner, the two parts being finally brought to a match—that is, to an equality of brightness.

Some of the common types of photometric field are shown in *Fig.* 1. The light in parts

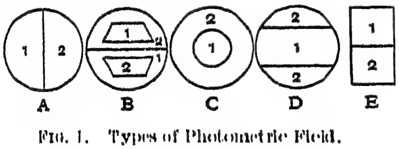


Fig. 1. Types of Photometric Field.

1 and 2 comes from beams 1 and 2, respectively. Type B is slightly different from the others and is known as a contrast field. Its construction is such that the trapezoidal parts are slightly darker than the surrounding field, equality of contrast being the criterion for a match. Type E is that usually obtained when a simple spectrometer with eyepiece is used, the light in parts 1 and 2 coming from different parts of the same slit. No one of these types has any special advantage over the others, unless it is type B, which is probably less fatiguing to the eye, and in the faint red and blue regions more sensitive than the others.

(ii.) *Conditions for High Sensibility.*—One of the most important conditions for obtaining maximum sensibility is that the dividing line or lines between parts 1 and 2 of the photometric field should be as fine as possible and should disappear when the match has been made. Another essential condition is that the field be of the proper angular size. A field subtending an angle of 3 to 5 degrees is considered very suitable. These two conditions are, of course, understood and, so far as possible, complied with by the makers of commercial instruments.

A third condition for obtaining maximum sensibility is that the field be of the proper brightness. In spectrophotometry one is seldom troubled by having too much light. For the sake of the measurements in the low-visibility parts of the spectrum—the red, blue, and violet—one should, in general, strive to have as much light as possible available. If the field becomes too bright throughout the green and yellow, it may easily be reduced by narrowing the slits of the instrument or by means of rotating sectors. It is also very important for accurate work that each separate part of the field be of uniform brightness.

In addition, the eye should be protected from all stray light; the room should be dimly lighted, if at all; and if the observer has to record his own observations, the illumination of the record sheet should be as low as can be used. A single glance at a bright sheet of paper or other light may incapacitate an observer's eye for several minutes if he is working at low brightness.

With any instrument or method, too, there will be certain conditions under which measurements may be made with maximum precision; and every instrument should be studied and used with this in mind. For example, at what value on the scale (*i.e.* at what slit-width, at what distance, at what angle of the nicol, etc.), at what concentration and thickness of solution, etc., will the percentage error in the final result be the least if an error of one division is made in the setting? Such a study is too often neglected.

(iii.) *The Sources of Illumination.*—When the relative spectral distribution of radiant power of a light is to be obtained, the following essential conditions should be met:

(*a*) A source of illumination must be available whose relative spectral radiant power is accurately known, so that it may be used as a standard.

(*b*) Light from this standard and from the unknown source must be subject to the same conditions; that is, they must travel along identical paths and be compared in turn with the same illumination, supplied by a third source.

(*c*) All the sources of illumination must be kept of constant value while measurements are being made.

If the spectral transmissive or reflective

[1] *Phys. Rev.* (2), 1919, xiv. 261.

properties of materials are to be studied, the apparatus should be so constructed, if possible, that the following conditions are true :

(*a*) The two beams of light should originate from the same source, so that fluctuation will not introduce errors.

(*b*) It should be possible to obtain diffuse spectral reflection as quickly and as accurately as spectral transmissive properties.

(*c*) In work on transmissive properties no error should be introduced or correction needed because of the change in length of the optical path caused by the beam traversing the specimen.

(*d*) In reflection work the illumination of the surface should be as nearly as possible equally diffuse from all directions, and the sample should be studied from a direction normal to its surface.

(*e*) The illumination should be intense enough to permit of very narrow slits being used; for the narrower the slit the purer the spectrum and the more accurate will be the measurements. An intense source of light will also make possible further and more accurate readings in the red and violet.

(*f*) In addition to heterogeneous illumination, sources of homogeneous illumination should be available at as many different wavelengths as possible, that all slit-width errors may be completely eliminated and the heterogeneous readings checked.

(iv.) *Methods of varying the Brightness of the Photometric Field.*—Different well-known methods are available for varying the amount of light which comes to form the parts of the photometric field. In spectrophotometry the actual amount of light which enters the slits of the instrument is not required; it is rather the relative amounts under specified conditions.

One may vary the brightness by varying the distance of a small source of light from the slit, the amount of light entering the slit for any two positions of the source being inversely proportional to the squares of the distances between the source and slit (or screen before the slit) in the two cases. This relation is very accurate, provided the diameter of the source is very small compared to the distances being used, *e.g.* 1/25.

A second accurate method is to use a pair of polarising nicol prisms in the path of the light. The light transmitted is then zero when the nicols are crossed and a maximum at a position 90 degrees from that, and the amount transmitted at intermediate positions is proportional to the square of the sine of the angle measured from the position of extinction. The faces of the nicols should be perpendicular to the direction of propagation of the light, the light entering the first nicol should be unpolarised, and the first nicol only should be rotated. Other combinations of polarising prisms may be also used, such, for example, as three nicol prisms (the middle one being rotated), or a Wollaston prism and nicol.

Varying the widths of the slits is a method often used with certain types of instruments, the brightness being taken proportional to the slit-width. This method is, however, "bad at heart," and should not be used where other methods are possible. In order to secure a suitable range of variation the slit-width has to be varied from a large part of a millimetre, which includes too much of the spectrum, making it very impure, down to a few hundredths of a millimetre where errors due to diffraction from the edges of the slit may become serious. It is probable, however, that this diffraction error is approximately eliminated by subtracting 0·02 mm. from all readings.[1] The error resulting from the use of too wide a slit may become enormous if, for example, one is measuring the transmission of substances like didymium glass or the solutions of certain rare earths which have very strong and narrow absorption bands. Methods for making the slit-width correction have been derived.[2] If it is necessary to use the slit-width method of varying the brightness, a thorough calibration should be made with a rotating sector or other means under the same conditions of wave-length and brightness as are to be used in the investigation.

A fourth method of varying the brightness is by the use of rotating sectors at speeds where flicker is no longer noticeable. The transmission of the sector as judged by the eye is accurately proportional to the aperture being used. Where this cannot be changed while the sector is rotating, the method is of greatest use as an auxiliary to increase the range of other methods. Sectors whose aperture may be changed while in operation have, however, been designed,[3] and in this form the method is as accurate and convenient as any other.

Other methods are sometimes used, such, for example, as interposing an absorbing wedge of varying thickness in the path of the light · but these are, as a rule, not as suitable as the others mentioned.

(v.) *Calibration of the Dispersing Prism.*—In spectrophotometric work the wave-length (or frequency) calibration need not be made as exactly as is required for most spectroscopic work. Because the amount of the spectrum being examined at any wave-length is usually one or more millimicrons, it

[1] Nichols and Merritt, *Phys. Rev.*, 1910, xxxi. 502.

[2] *Ibid.* xxx. 328; Hyde, *Astrop. Jour.*, 1912, xxxv. 237.

[3] Hyde, *Astrop. Jour.*, 1912, xxxv. 257; Abney, *Researches in Color Vision*, p. 68; Brodhun, *Zeitschr. f. Instrk.*, 1907, xxvii. 8. The instrument described in the last reference has a rotating beam of light and a stationary but variable sector.

is as a rule sufficient if the calibration can be guaranteed to a few tenths of a millimicron. The use of symmetrically opening slits is, of course, highly desirable, so that changing the width of a slit will change the effective wave-length as little as possible.

The usual method of calibration is to locate the position of a number of well-known spectral lines on the arbitrary scale of the instrument. (If the scale is already graduated in wave-lengths it should be checked very carefully before using.) These lines should be well distributed from the extreme red to the extreme violet, no extensive part of the spectrum being omitted. As a rule, the more lines one observes, the more accurate will be the calibration. The sources often used for this purpose include the mercury, helium, and hydrogen vacuum lamps, and the electric arc of metals like cadmium, copper, zinc, etc., whose spectra contain not too many lines.

The known values of wave-length, or frequency, and the corresponding readings on the arbitrary scale of the spectrophotometer should be plotted to a proper scale and a smooth curve carefully drawn through these points. It is then a simple matter to obtain the value of the arbitrary-scale reading for any desired wave-length or frequency. The frequency curve will be found to have much less curvature, and is therefore safer to use if but few known values are available. Empirical equations have been found which represent the wave-length curve quite accurately and which may be obtained when three or four lines well distributed over the range to be used have been accurately located. The constants of the equation having been obtained from these values, it may then be used to compute the value of the arbitrary-scale reading for any desired wave-length. The use of equations, however, is apt to be quite laborious, and the graphical method first mentioned is ordinarily sufficient.

§ (7) Individual Types. (i.) *The Lummer-Brodhun Spectrophotometer.*—The Lummer-Brodhun spectrophotometer [1] is a well-known instrument, a brief description of which will illustrate some of the points already mentioned. (A much more complete illustration of the various principles underlying the use of any spectrophotometer will be found in Part IV.) An outline diagram (plan) of the instrument is shown in *Fig.* 2. The light in beams 1 and 2 enters slits S_1 and S_2 respectively of collimators T_1 and T_2, and is made parallel by lenses L_1 and L_2. These two beams of parallel light then enter the Lummer-Brodhun cube at right angles to each other.

This cube is composed of two pieces of glass. Part 2 is a right-angled total-reflection prism, and if used alone would reflect all of beam 2 towards P and all of beam 1 away from it. A small part of the surface of part 1 is, however, cemented on to part 2 with Canada

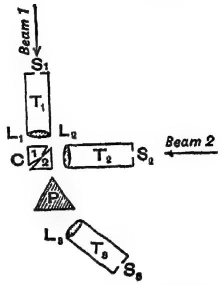


Fig. 2.—The Lummer-Brodhun Spectrophotometer.

T_1, T_2 = collimating telescopes; T_3 = observing telescope; S_1, S_2, S_3 = vertical slits; L_1, L_2, L_3 = lenses; P = dispersing prism; C = Lummer-Brodhun cube—giving field of type C, *Fig.* 1.

balsam, the rest of that surface of part 1 being ground away so as not to be in contact with part 2. Under these conditions the central part of beam 1 and the outer part of beam 2 will form the photometric field of type C, *Fig.* 1. (Type B is another common form of photometric field found with Lummer-Brodhun spectrophotometers.) This is viewed by the eye placed at slit S_3, no eyepiece being used. Uniformity of the field is obtained by using pieces of ground glass before slits S_1 and S_2 or by having the light come from illuminated pieces of magnesium carbonate.

This instrument is especially well adapted to compare the relative spectral distribution of radiant power of two light sources. When this is being done, it is best to compare each one of them in turn with a third source, this third source being placed, for example, before S_1, giving rise to beam 1, and the other two being in turn placed before S_2, giving rise to beam 2. Any one of the methods already indicated, or a combination of them, may be used to vary the amount of light entering slits S_1 and S_2 and thus bring the two parts of the photometric field to equality.

In transmission work two sources may be used—or better, only one, the light being reflected by mirrors or pieces of magnesium carbonate into the two slits. This will depend somewhat upon the method used for varying the brightness to obtain the match when the sample being examined is placed in or out of one of the beams of light. For the measurement of diffuse or specular spectral reflection special auxiliary illumination apparatus is necessary. This may be easily devised.

(ii.) *The König-Martens Spectrophotometer.*—A second well-known instrument operating on a very different principle is the König-Martens

[1] *Zeitschr. f. Instrk.*, 1892, xii. 132. See "Photometry and Illumination," § (18).

spectrophotometer.[1] In this case variation of brightness of the photometric field is brought about within the instrument by polarising the light and intercepting it with a nicol prism. An outline diagram (plan) is given in *Fig.* 3.

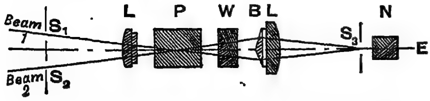


FIG. 3.—The König-Martens Spectrophotometer.

L = lenses; W = Wollaston prism; B = bi-prism, giving field of type A, Fig. 1; P = dispersing prism (the deviation is away from the plane of the paper); N = nicol prism; S_1, S_2, S_3 = horizontal slits (S_1 and S_2 are parts of the same slit); E = position of the eye.

The light from beams 1 and 2 traverses the instrument approximately as shown. By means of the Wollaston prism each beam is split up into two parts vibrating at right angles to each other. One part of each beam is lost within the instrument; the remaining parts, polarised mutually perpendicular, are brought by means of the bi-prism and lens to form the two halves of a photometric field, which is viewed by the eye through the nicol prism. Errors do not result from the partial polarisation by the dispersing prism P, because the light is polarised by W after leaving P. The field is of type A, *Fig.* 1, the dividing line being produced by the apex of the bi-prism. Uniformity of field may be secured in the same way as with the Lummer-Brodhun spectrophotometer.

Since the two beams of light reaching the nicol are polarised in directions at right angles to each other, a rotation of the nicol prism will obviously increase the brightness of one half of the field while decreasing that of the other, one becoming a maximum when the other becomes extinct, and *vice versa.* The angular scale is made to read zero for one of the positions of extinction.

If one of the original beams, *e.g.* beam 1, is kept constant while the other, beam 2, is varied, and a match has been obtained in two different cases, the brightness of beam 2 in the first case to that in the second case will be given by the ratio of the squares of the tangents (or cotangents) of the angles of match. If beams 1 and 2 may be interchanged, the ratio of brightness is the product of the tangent of the angle of match in one case by the cotangent in the other. Methods of using this instrument are explained in detail in Part IV.

(iii.) *Other Instruments.* — These two examples are illustrative of the great variety in method which has been employed to bring two beams of light of the same wave-length into the proper juxtaposition in a photometric field, and to enable either or both beams to be varied in brightness at will in a continuous and known manner. They are illustrative of the two general classes into which visual spectrophotometers are often divided—viz. polarisation instruments and others (mostly of the variable-slit type). They are also illustrative of another general division which might be made—viz. those which are designed primarily for the comparison of the spectral distribution of radiant power in two sources, and those in which the measurement of the spectral transmissive properties of materials is the main purpose in mind. With the proper auxiliary apparatus, however, most instruments may be used for both these purposes as well as for the measurement of diffuse reflection.

There are many other well-known instruments in addition to the two already described. One of the earliest types was the double-slit spectrometer of Vierordt,[2] in which a single collimator carried a divided slit, either half of which could be varied in width independently of the other, values being computed from these known slit-widths. The type of field is that represented by E of *Fig.* 1. The accuracy possible with this instrument was increased by later investigators. Krüss [3] used slits which opened symmetrically on either side instead of having one side fixed and the other variable. Further improvements in this method by various investigators consisted in devising means for making the dividing line of the photometric field as fine as possible and at the same time separating the two beams of light a short distance to make it easier to insert a specimen in one beam without interfering with the other beam. One of the modern varieties of the double-slit instrument, but with two collimators and a Lummer-Brodhun cube, is called the Differential Spectrophotometer,[4] in which the two slits work by a single screw in such a way that the sum of the two slit-widths remains constant.

The Brace [5] spectrophotometer is very similar in action to the Lummer-Brodhun, it having two collimators widely separated. In this instrument, however, the photometric field is formed in the prism itself; it is of type D, *Fig.* 1.

While much valuable work has been done with variable-slit instruments, the general accuracy and reliability of all of them are increased if some other means of varying the brightness of the photometric field is used; and this has often been done.[6]

[1] *Ann. der Phys.* (4), 1903, xii. 984.

[2] *Pogg. Ann.*, 1869, cxxxvii. 200; 1870, cxl. 172.

[3] *Rep. d. Phys.*, xiii. 267; *Zeitschr. f. analytische Chemie*, 1882, xxi. 182.

[4] Shook, *Astrop. Jour.*, 1917, lxvi. 305.

[5] *Phil. Mag.* (5), 1899, xlviii. 420; *Astrop. Jour.*, 1900, xi. 6.

[6] *E.g.* Lemon, *Astrop. Jour.*, 1914, xxxix. 204.

Spectrophotometers which make use of polarising apparatus to vary the brightness of the field do not require the use of variable slits, but retain the same ease and speed of making settings. They should, however, likewise be subjected to thorough test as to the possible presence of errors resulting from the polarising action of the dispersing prism or other parts of the apparatus. The objection often made to polarisation instruments, viz. the loss of light within the instrument, has become of little weight because of the high efficiency, especially in the blue and violet, of the gas-filled incandescent lamps now available. In addition to the König-Martens instrument, elsewhere described in detail as to construction and use, the following methods have been used.

In Glazebrook's[1] instrument two widely separated beams are each polarised by a nicol prism, but mutually perpendicular. These are brought into juxtaposition and viewed by the eye through a third nicol. By rotating this nicol a photometric match is obtained, values being computed by the $\tan^2$ relation as with the König-Martens instrument.

In another type of instrument,[2] similar in many respects to the König-Martens spectrophotometer, a single collimator is used. Light from two separated parts of the slit are polarised mutually perpendicular by means of a Rochon prism. The two beams, after passing through a nicol prism, finally form two contiguous spectra, any width of which may be examined by the eye. Equality of brightness is obtained, again, by rotating the nicol and values computed from the $\tan^2$ relation.

A polarisation instrument of different type is that by Crova,[3] in which the light in only one beam is polarised. In that beam two nicol prisms are placed and equality of brightness obtained by rotating one of the nicols, the $\sin^2$ relation being used for the computation. In this instrument the two beams of light come from different parts of the same slit, but that in one half is deflected in by means of a total reflection prism, its original direction being at right angles to the other. Three nicols instead of two have sometimes been used.[4] In this case the middle one is rotated; and the $\sin^4$ relation must be used in computations.

In Hüfner's spectrophotometer[5] one beam only is polarised, one nicol being placed in one beam before the collimator slit, the other in the observing telescope. Hüfner's rhombic-prism scheme is used to bring two slightly separated beams into the proper juxtaposition through different parts of the same slit.

The action of a polarisation instrument by Wild[6] depends upon the vanishing of the interference bands formed by the proper superposition of the two beams of light from any one source.

Various devices are available by which any spectrometer may be converted into a spectrophotometer; in fact, some of the instruments already mentioned are practically that. Total reflection or rhombic prisms are used for this purpose, bringing two beams of light into separate but contiguous parts of the collimator slit. Auxiliary means must be provided for varying the brightness of one or both beams. Such methods have been used by Nichols,[7] Houston,[8] Ives,[9] Nutting,[10] and others.

One of the most recent forms of spectrophotometer is the Keuffel and Esser Color Analyser, a direct-reading instrument. A constant-deviation spectrometer is used. A bi-prism placed over the telescope lens gives a field of type A, *Fig.* 1, with horizontal dividing line. The two beams come from different parts of the same slit. The brightness of one beam relative to the other is varied by means of a variable rotating sector, and the scale reads transmissive or reflective percentages directly.

A somewhat different method from any of the above is that by Abney,[11] where the illumination of two shadows, side by side and touching one another and cast on a screen by placing a rod in the path of the two sets of rays, is equalised by placing a variable-aperture rotating sector in the path of one of the beams.

Mention might also be made of spectrophotometers operating on the flicker principle.[12] This subject, however, is one for heterochromatic photometry rather than spectrophotometry, and is, therefore, untouched here.

§ (8) Photometers and Homogeneous Light.—Considerable work of value in the line of checking and studying other methods, or in actual experimental work, if a spectrophotometer is not available, may be done by means of a photometer, source of homogeneous light, and selective ray filters. The Martens photometer[13] has been used at the U.S. Bureau of Standards for this kind of work; other types would also be suitable. The mercury, hydrogen, and helium lamps are of value because of having several widely separated lines of sufficient brightness to be used. If

[1] *Camb. Proc.*, 1883, iv. 304.
[2] Glan, *Wied. Ann.*, 1877, i. 351.
[3] *Ann. de Chim. et Phys.* (5), 1883, xxix. 556.
[4] Zenker, *Zeitschr. f. Instk.*, 1883, iv. 83.
[5] *Zeitschr. f. physik. Chemie*, 1889, iii. 562.
[6] *Rep. d. Phys.*, 1883, xix. 812; *Wied. Ann.*, 1883, xx. 452.
[7] *Phys. Rev.*, 1894, ii. 138. Diffraction grating also used for dispersion.
[8] *Phil. Mag.*, 1908, xv. 282.
[9] *Phys. Rev.*, 1910, xxx. 446.
[10] *Bur. Stand. Sci. Pap.* No. 115, 1911; see also The Nutting Photometer, advertised by Hilger.
[11] *Researches in Colour Vision*, p. 74.
[12] *E.g.* Lummer-Pringsheim, *Jahresber. d. Schles. Ges. f. Vaterl. Kultur*, 1906; *Beibl.*, 1907, 466; Milne, *Proc. Roy. Soc.*, Edinburgh, 1912–13, xxxiii. 257.
[13] *Phys. Zeit.*, 1900, i. 299.

all three sources are available, the following wave-lengths may be obtained:

Hg—404·7 + 407·8, 435·8, 546·1, 576·9 + 579·1 mμ;
H —434·1, 486·2, 656·3 mμ;
He—587·6, 667·8 + 706·5 (706·5 weak in comparison to 667·8), 706·5 mμ.

The helium lines in the blue are too close together to be efficiently separated.

Glasses for filtering out these lines are described in *Bur. Stand. Tech. Pap.* No. 148; and gelatine filters for the same purpose are advertised by the Eastman Kodak Co., Rochester, N.Y.

A Variation-of-Thickness Photometer has recently been designed for measuring the transmissivity of liquids. The following are the essential features:[1]

"Two beams of light proceed horizontally, one above the other, from a uniformly illuminated vertical surface.

"A rotating sector disc of known transmission can be interposed in the upper beam.

"By means of two totally reflecting partially immersed rhombs, the lower beam is diverted through a variable thickness of oil determined by the distance between the rhombs, one of which is carried by a slide on a track parallel to the beams of light. The oil is contained in a horizontal trough with open top into which the rhombs dip.

"The two beams are brought into juxtaposition in the photometric field by means of an arrangement of biprisms.

"The thickness of oil in the lower beam is varied until its transmittance[2] equals the transmission of the sector disc as judged by equality of brightness in the photometric field.

"A suitable scale provides for *direct* reading of transmissivity[3] or the logarithm of transmissivity.

"Light of definite wave-length may be obtained either by:

"(1) Use of light filters with a *selective source*, such as a mercury lamp, or

"(2) Dispersion with a prism, which converts the instrument literally into a spectrophotometer."

IV. Description and Comparison of Methods

For several years the subject of spectrophotometry has been given considerable attention in the chromatics section of the U.S. National Bureau of Standards. This has been necessitated by the many requests continually being received there for tests of the spectral transmissive and reflective properties of all kinds of material, and by the large and increasing demands for the standardisation and specification of colours. Spectrophotometry, as already noted, is the fundamental basis of all work of this kind.

It was early realised that sufficient attention had never been paid to the detection and elimination of errors by the comparison and counterchecking of radically different methods. It is no trouble at all to obtain, for example, "smooth" spectral transmission curves by any one method, the observational error being perhaps very small, and repetition of the measurements being easy to obtain. It is only when the same specimen is measured by a different method over the same range that one realises there may be enormous errors present though previously undetected. Moreover, any one method is limited in range and becomes somewhat unreliable as the outer regions are approached. The limits of the visible spectrum are often taken as 400 and 750 mμ, but little accurate visual work has ever been done at wave-lengths less than 450 and greater than 700 mμ. And yet cases arise where measurements within these regions are of great importance.[4]

For these reasons several independent methods of spectrophotometry have been developed, in addition to the construction of much auxiliary apparatus for use with instruments already available; and by means of the counterchecking thus made possible the work has been brought to a high degree of efficiency and reliability. It has therefore been thought that the ends of this article can be served in no better way than by describing in considerable detail the apparatus and methods now being used there. Most of this material has been obtained from published papers, and reference will be made to these at appropriate places.

§ (9) The Visual Method with the König-Martens Spectrophotometer and Auxiliary Apparatus.—As already noted, a spectrophotometer for general purposes should be of such a kind and used in such a way that any one of three different kinds of experimental work may be accurately and conveniently performed with it; that is, it should be available for the determination of (i.) the relative spectral distribution of radiant power of the various sources of illumination in commercial use at the present time, (ii.) the spectral transmissive properties of all kinds of substances, up to a considerable thickness if necessary, and (iii.) the diffuse spectral reflection of a great variety of materials. The König-Martens spectrophotometer, when used in the proper way and with suitable auxiliary apparatus, meets these conditions admirably, as will be illustrated.

The usual working range of the instrument has been from about 450 to 700 mμ, with slit-widths of 0·10 mμ in collimator and telescope. But by widening the slits reliable

[1] Priest, *The Cotton Oil Press*, July 1920, p. 96.

[2] *Transmittance*: the ratio of the radiant power incident on the second surface of a substance to that transmitted by the first surface.

[3] *Transmissivity*, t: Let T = transmittance, b = thickness; then $t = \sqrt[b]{T}$.

[4] *E.g.* in chromatic camouflage, *Jour. Opt. Soc. America*, 1920, iv. 390.

work has been done as far as 420 and 750 mμ with the heterogeneous source. The brilliant mercury line at 435·8 mμ has been of great assistance in obtaining reliable readings in this region. The use of the bright homogeneous lines of the mercury and helium lamps has proved of considerable value for checking purposes.

Methods of computation, which would ordinarily be very laborious with this instrument, have been simplified by the design of slide-rules, by which the various computations involving tangents and cotangents, or their squares, may be performed as conveniently and accurately as in ordinary numerical work with the ordinary slide-rule.

(i.) *Comparison of Light Sources.*—The use of this spectrophotometer for the comparison of light sources is described in a paper on the "Colour and Spectral Composition of Certain High-Intensity Searchlight Arcs."[1] By the method therein given in detail it is possible to obtain the relative spectral distribution of radiant power of even a varying source like the electric arc. The following diagram and description of the method are taken from that paper.

An illumination box was constructed and fastened to the spectrophotometer immediately in front of the slits (*Fig.* 4). This consisted

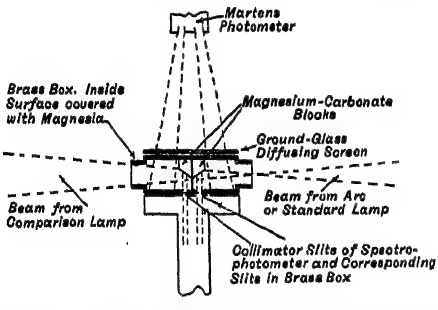


FIG. 4.—Details of Illumination Box, used with the König-Martens Spectrophotometer for the comparison of light sources.

of a brass box divided into two compartments by a brass partition. "The interior of each compartment was coated white with magnesium oxide, put on by burning magnesium ribbon, and blocks of magnesium carbonate were placed as shown. Thus the whole interior was white and the reflection from all surfaces was entirely diffuse. The magnesium-carbonate blocks were illuminated as shown, and light from these two illuminated surfaces entered the double slit of the spectrophotometer through small openings in the illumination box. The whole interior of the box is, of course, illuminated by diffuse reflection from the magnesium-carbonate blocks, and two beams of light from the sides of the box next to the spectrophotometer pass through openings in the opposite side and are compared by means of the Martens photometer." A red glass filter was placed over the eyepiece of this photometer to eliminate the colour difference between the arc and the comparison lamp, and a ground-glass diffusing screen used, as shown, to make the photometric field still more uniform and of the proper brightness. Lenses were used to obtain sufficient illumination of the magnesium-carbonate blocks, care being taken that this illumination was uniform, and that chromatic effects were avoided. Rotating sectors of known aperture were used on one side or the other, as needed, so that the comparison field of the spectrophotometer might be brought to equality within a suitable region on the scale.

[1] *Bur. Stand. Tech. Pap.* No. 168, 1920.

The substitution method of measurement should always be used in this kind of work; that is, a comparison lamp should be kept on one side and light from it compared first with light from a standard lamp whose relative spectral radiant power has been determined radiometrically, and then with that from the lamp, arc, or other source of illumination whose relative spectral radiant power is to be determined. If the unknown source of illumination is one that may be easily maintained constant, *e.g.* an incandescent light, but two observers are needed—one to watch and control (by means of a potentiometer) the voltage around the lamps, thus maintaining constant illumination on either side, the other to make observations at different regions in the spectrum. If the unknown source is an arc (in this case the Sperry searchlight arc), more observers will be required.

"Five people were necessary in making the observations on the arc, one to watch and control the arc, recording the readings of the voltmeter; a second to regulate the current through the arc, recording the readings of the ammeter; a third to observe the illumination through the Martens photometer; a fourth to take readings on the spectrophotometer; and a fifth to record these readings and keep the proper voltages on the lamps by means of the potentiometer. The procedure for taking a series of observations under these conditions was as follows:

"The arc was allowed to run a few minutes in order that conditions might become as steady as possible. While this was being done, the comparison lamp was kept at its proper voltage and the variations of the illumination from the arc followed on the Martens photometer. Finally, the photometer was set for equality of brightness between the two sources at what seemed to be the best mean value of the illumination from the arc. This setting was then unchanged during the run. Now, for any given wave-length, the variations in the illumination from the arc were followed on the spectrophotometer, the attempt being made to keep the two halves of the

photometric field always equal in brightness. Whenever conditions were right, as shown by the Martens photometer, a sharp signal was given and the reading of the spectrophotometer taken at that instant. It was not attempted to do this, of course, if the illumination varied too rapidly through the proper value. Usually the variations were quite slow, the illumination often staying constant at the correct value for a considerable part of a minute. It was noticed during the course of the investigation that the value of the current through the arc would have served nearly as well as the auxiliary photometer for enabling measurements to be made always at the same illumination; for the current was practically always at a given value, about 75 amperes, when the signal was given by the observer at the Martens photometer.

"At each wave-length from four to ten readings were taken on the spectrophotometer, the number depending on the agreement obtained among the readings. Measurements were made between wave-lengths 430 or 440 and 710 mμ usually at every 20 mμ.

"After this series of measurements was completed, the total reflection prism was set up to locate the position of the arc and kept there until the standard lamp had been put in position. It was then removed, and another series of measurements made at the same wave-lengths as previously on the arc. Only two observers were now necessary, one to make the measurements on the spectrophotometer, the other to record the data and keep the voltages constant with the potentiometer.

"In this manner all the data were obtained, either a complete run or check points being taken on the standard lamp between any two of the runs on the arc.

"The spectral distribution of radiant power of the arc was computed as follows: All the values of the angles of the nicol prism read on the spectrophotometer for the standard lamp were plotted at the proper wave-lengths and a smooth average curve drawn through these points. All values for the standard lamp used in the computations were taken from this curve. Thus the same values of the standard lamp were used for all computations. For each wave-length the square of the cotangent of the average angle, read on the spectrophotometer when the arc was in position, was divided by the square of the cotangent of the angle as read from the curve when the standard lamp was in position. When this is done for all the wave-lengths, the radiant power of the arc relative to that of the standard lamp is obtained. By multiplying these values by the known relative values of radiant power of the standard lamp at the different wave-lengths, the relative spectral distribution of radiant power in the arc was obtained."

(ii.) *Spectral Transmissive and Reflective Measurements.*—No complete account of the illumination apparatus which has been developed for use with the König-Martens spectrophotometer has yet been published, but brief reference to it has been made in two papers.[1] The essential features are as follows:

A metallic box was constructed large enough to contain ten 500-watt, incandescent, Mazda-C lamps and three large M-shaped mercury lamps well distributed throughout the interior. A system of air and water cooling prevents undue heating of any part of the apparatus. The interior of the box was painted white and finally coated with magnesium oxide, giving diffuse illumination from all parts.

The box was set up in front of the spectrophotometer at sufficient distance to allow samples of considerable thickness to be interposed. In the side of the box next to the slits two small openings were cut just large enough to include the angles subtended by the slits and lens of the spectrophotometer. On the opposite side of the box, two pieces of magnesium carbonate are placed, in such a position as to cover the same angles. These two surfaces of magnesium carbonate are illuminated from all directions within the box by means of the lamps and the white walls, the incandescent lamps or the mercury lamps being used at will.

A telescope arrangement within the box prevents any of the light entering the slits of the spectrophotometer except that coming from the surfaces of magnesium carbonate. Thus the comparison field of the instrument is filled with light, half from one magnesium-carbonate surface, half from the other; and any variations in the amount of light from the lamps will be exactly neutralised. The lamps may thus be run on an ordinary, fluctuating circuit, even the flickering of the mercury lamps causing no trouble. Throughout the part of the spectrum of high visibility only a part of the lamps need be used, and these can be run below the maximum efficiency to prolong their life. It is only in the red, blue, and violet that it is necessary to make use of all available light.

When the spectral transmission of any material is to be obtained, the sample is placed in its carrier between the box and the slits and made to intercept the light first in one beam and then in the other. This interchange may be quickly effected by the observer without leaving his seat. For any one position in the spectrum he first takes from one to five readings with the sample in one beam; he then transfers the sample to the other side and takes twice as many readings; finally he returns it to the first beam and repeats his readings there to detect any possible change of conditions. The transmission is then computed by the product of the tangent of one set of readings with the cotangent of the other.

When the spectral diffuse reflection of any

[1] *Bur. Stand. Tech. Pap.* No. 148, 1920; No. 167, 1920.

material is to be measured, a sample is cut to the right size and substituted for one of the magnesium-carbonate surfaces. Interchange of sample and the other magnesium-carbonate surface is easily made by the observer, and a set of readings taken analogous to those already described for transmission work. The result will be the reflection relative to that of magnesium carbonate. This latter is nearly 100 per cent.

The carriers for holding the specimens in both transmission and reflection work were very carefully made so that in transferring the specimen from one beam to the other it might always be returned to exactly the same position. This is very important, for it often happens, especially in reflection measurements, that the samples are not absolutely uniform throughout. The carrier for transmission measurements has, of course, to be adjustable in order that specimens of varying thickness and diameter may be accommodated.

In connection with the study of the spectral transmissive properties of dyes and other solutions, special apparatus has been constructed. The cells to hold the solution and solvent were made after the pattern of those put out by Hilger for use with the sector photometer. While it is unnecessary to describe this cell in detail, it has four valuable characteristics which should be possessed by all cells used in this kind of work: (1) evaporation is prevented; (2) nothing but glass (or quartz) comes in contact with the solution; (3) no wax or cement of any sort is used; (4) in order to refill, it has to be completely taken apart, and may therefore be easily and thoroughly cleaned for each new solution.

Inasmuch as some solutions change their transmissive properties rapidly with change in temperature, holders for these cells have been constructed, in which water at constant temperature is circulated, the constant temperature and circulation being maintained by a pump-and-thermostat arrangement. The solutions may thus be kept constant within a fraction of a degree at any desired temperature over a considerable range. A detailed description of this auxiliary apparatus for use with solutions has been published.[1] Similar apparatus has also been constructed for use with the other methods whose description follows.

§ (10) THE PHOTOELECTRIC NULL METHOD. (i.) *Advantages of the Method.*—The well-known difficulty of obtaining accurate spectrophotometric measurements in the blue and violet by any other method led to the development of the photoelectric null method herein described.[2] The potassium-hydride, gas-filled, photoelectric cells now on the market make it possible to obtain as reliable determinations throughout the blue and violet, and even beyond those regions, as are obtained at other wave-lengths by other methods.

The photoelectric cell has been used in various ways by different investigators.[3] Either the galvanometer or electrometer may be used with it; and in practically all measurements of spectral transmissive or reflective properties some form of deflection method has been used, much valuable work having been performed. For example, the absolute specular spectral reflections of a large number of metals have been determined in the ultra-violet between wave-lengths 380 and 180 $m\mu$.[4]

In various ways, however, a null method seems to be more suitable than a deflection method. It has been concluded by the majority of investigators that it is not safe to assume a direct proportionality between photoelectric current and exciting radiant power; and therefore any cell used in a deflection method, unless it be a method of equal deflections, must be tested and calibrated if there is deviation from this straight-line relationship. The so-called dark current must also be eliminated or a correction made, unless the radiant powers used are so great that it may be neglected, which is seldom the case in spectral work.

When the electrometer is used, still further difficulties are met by one using a deflection method. A strict proportionality between deflections of the disc, as read by means of the mirror and scale, and the potential acquired by the quadrants, must be proved by test or a calibration made. Timing devices are also necessary with their accompanying inconvenience and possible errors. But the electrometer is so much more sensitive than the galvanometer that it is desirable to use it if possible.

In the following description of the null method are shown the ways in which the questions of the current-irradiation[5] relationship and the dark current may be avoided and their errors eliminated while retaining the extreme sensitivity of the unshunted electrometer.

(ii.) *Spectral Transmissive Measurements.*—The principles on which the null method is based are very simple and may be understood in connection with the accompanying diagram (*Fig.* 5).

"When the earth connection is made at U, both pairs of quadrants are at zero potential. Under this condition the charged electrometer needle will be at rest. Let the voltage applied to P_1 be V_1 and that applied to P_2 be V_2, and let the total resistances (mainly that of

[1] *Bur. Stand. Sci. Pap.* No. 440, 1922: "I. The Seven Food Dyes."

[2] Gibson, *Jour. Opt. Soc. Am.*, Jan.-Mar. 1919; *Bur. Stand. Sci. Pap.* No. 349, 1919.

[3] Résumé, Coblentz, *Bur. Stand. Sci. Pap.* No. 319, 1918.

[4] Hulburt, *Astrop. Jour.*, 1915, xlii. 205.

[5] The terms irradiate and irradiation, as emphasised by Ives (*Astrop. Jour.*, 1917, xlv. 39), should be used analogously to the terms illuminate and illumination when radiant energy rather than light is discussed.

the photoelectric cells) between U and G′ be R_1 by way of P_1 and R_2 by way of P_2. Then through P_1 a current will flow of magnitude $I_1=V_1/R_1$ and through P_2 of magnitude

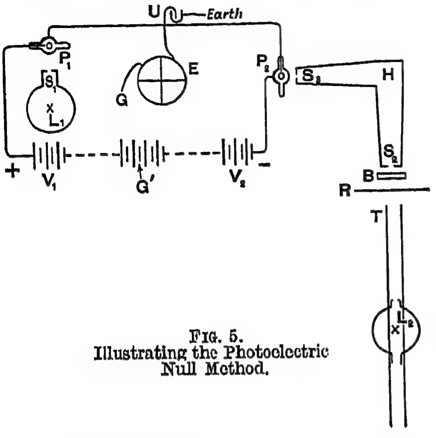


FIG. 5.
Illustrating the Photoelectric Null Method.

E = Dolezalek quadrant electrometer, disc charged to +150 volts; motion shown by spot of light and scale, as usual; P_1P_2 = Kunz photoelectric cells; GG′ = earth connection, zero potential; U = make-and-break mercury connection; V_1V_2 = voltages applied to P_1 and P_2, respectively; H = position of Hilger constant-deviation glass-prism spectrometer with slits S_2 and S_3 ($S_2=S_3$ always); B = location of specimen whose transmission is to be measured (means provided for moving in and out of beam); R = rotating sector to increase range of measurements; T = track on which distance of 600-watt lamp L_2 from slit S_2 may be varied; S_1 = slit by which irradiation of P_1 by miniature lamp L_1 may be varied, the amount being accurately proportional to the width of S_1.

$I_2=V_2/R_2$. No radiant energy to which the photoelectric cells are sensitive is considered as falling on them, I_1 and I_2 being what are ordinarily called the dark currents.

"If the earth connection at U is broken, the electrometer disc will be deflected (shown by the drift of the spot of light) unless I_1 is exactly equal to I_2. If they are not equal, they may be made so by changing the relative values of V_1 and V_2. This is done by varying the point of ground connection at G′.

"Now let P_1 and P_2 be irradiated by L_1 and L_2. This will bring about a great increase in the photoelectric currents I_1 and I_2, and the irradiations may be so adjusted (by varying the currents through the lamps, the width of S_1, or the distance of L_2) as to make I_1 again exactly equal to I_2, as shown by the zero motion of the electrometer disc when the ground connection at U is broken. Now let the irradiation of P_2 from L_2 be reduced by interposing the specimen B. Then I_2 will be less than I_1, and, with U broken, the spot of light will be deflected. Two methods may be employed to make I_1 and I_2 again equal, a means of obtaining a measure of the transmission of B being thus possible: (1) The amount of radiant energy falling on P_1 may be decreased by narrowing S_1 until I_1 becomes equal to I_2, the ratio of the slit-widths in the two cases being a measure of the transmission; or (2) the amount of radiant energy falling on P_2 may be increased by moving L_2 nearer the slit S_2 until I_2 again becomes equal to I_1, the inverse ratio of the squares of the distances in the two cases giving the transmission.

"The value of the transmission obtained by the two methods will be the same only in case the dark currents are exactly equal and in case the two photoelectric cells obey exactly the same irradiation-current law. In case either of these conditions is not fulfilled, the value of the transmission obtained by the first method will be in error, but by the second method the accuracy of the value of the transmission will be unimpaired though either or both of these conditions are not fulfilled. Therefore, the transmission is measured by merely varying the distance of L_2 from the slit S_2, this distance being the only variable in the operation; for the width of the slit S_1, the amount and quality of the radiant energy falling on P_1 and P_2, the photoelectric currents I_1 and I_2, and the currents through L_1 and L_2 are unchanged, whether the specimen is in or out."

"The distance of L_2 from S_2 can never be made less than 46 cm. because of the other apparatus on the bench, such as the sector, carrier, etc. This, combined with the fact that the filament of L_2 is in a plane only 1 cm. square, enables the inverse-square law to be assumed; that is, the amount of radiant energy entering the slit S_2 varies inversely as the square of the distance of L_2 from S_2. The rotating sector serves as a means of making sure that the apparatus keeps in perfect working condition from day to day. The length of the photometer bench enables L_2 to be moved back to 255 cm. from the slit. Therefore, the range of transmission possible with the 10 per cent sector is from 1·00 to approximately 0·004."

Between 410 and 550 mμ inclusive, the method is very reliable as shown by agreements with other methods. These will be discussed later. If a quartz-prism spectrometer were used, measurements could be extended into the ultra-violet. In the long wave-length direction the cell itself soon becomes very insensitive. The usual working range of the apparatus has been from 390 to 600 mμ inclusive, but values have been obtained as far as 380 and 650 mμ.

(iii.) *The Measurement of Diffuse Spectral Reflection and other Applications.*—The apparatus was designed primarily for the measurement of spectral transmissive properties, but has also been used to obtain the diffuse spectral reflection of substances relative to that of magnesium carbonate. For this purpose the lamp L_2 is moved forward, and by means of lens and mirror the rays are brought to a focus upon the substance to be studied, the angle of

incidence being approximately 45 degrees. The diffuse reflection leaving this substance (or the magnesium carbonate) at right angles then enters the slit S_2, and the resulting photoelectric current in P_2 is balanced by that in P_1 excited by L_1, as was done in the case of spectral transmission. The method of obtaining the balance now, however, has to be by the means of varying the width of S_1.

The method has not been as thoroughly tested as that for the transmission measurements. Comparison with other methods, however, has shown that for approximately diffusing surfaces the method is very reliable except at wave-lengths where the sensitivity is low or the reflection being studied is very small in value. Where values have to be obtained by varying the width of S_1, however, the accuracy is about half that resulting when the distance of L_2 from S_2 can be varied as explained under spectral transmission. But as already noted this accuracy is dependent upon the two cells obeying the same current-irradiation law.

Measurements of relative specular reflection have also been made by the method of varying the width of S_1.[1]

"The apparatus is well adapted for the comparison of spectral distribution of radiant power of two sources over the same range of wave-lengths as used in transmission and reflection measurements. If the two sources are such that they obey the inverse-square law, the method of varying their distances from S_2 could be used, S_1 being kept constant. If the inverse-square law were not obeyed by either source, the other method could be used, their distances from S_2 being kept constant and S_1 varied to obtain a balance."

(iv.) *The Elimination of Errors.*—In setting up the apparatus the usual precautions were taken to eliminate so far as possible all mechanical vibration, moist air, stray radiant energy, and electrical disturbances. This was accomplished by the proper use of a metallic enclosure for the electrometer and photoelectric cells, and by resting this upon a cement pier in a basement room.

The use of the null method eliminates the errors liable to be found with electrometer-deflection methods. "Such errors may be those due to the deflections of the electrometer disc as given by the spot of light on the scale not being strictly proportional to the charge or potential acquired by the quadrants, those due to leakage of charges because of imperfect insulation or moist air, or those connected with the use of timing devices."

The method of varying the distance of L_2 from S_2, as illustrated under spectral transmission measurements, eliminates two other possible errors: (1) "Errors due to the photoelectric current not being strictly proportional to the radiant power incident on the cell. No calibrations of the cells are necessary, and it makes no difference what the relation is between radiant power and photoelectric current;" (2) "Errors due to what is ordinarily known as the dark current. It makes no difference whether or not it is eliminated. Nor do the dark currents through the two cells have to be the same, though approximately this condition is desirable for convenience' sake."

Errors which might result from undetected changes of the dark currents through P_1 or P_2 or of the currents through L_1 or L_2 are largely eliminated by taking the first and third readings with the sample in the beam, the second with it removed, etc.

Errors arising because of the inverse-square law not being exactly obeyed by L_2 "would be expected to be very small because of the filament occupying an area of only 1 cm.^{2}, and because it is not possible to move the filament closer to the slit than 46 cm. It has been tested by means of the rotating sector. Other investigators[2] have found that Talbot's law holds for the photoelectric cell; and a great number of observations at different times and at many wave-lengths on the apparatus herein described prove that the inverse-square law is obeyed, assuming Talbot's law to hold, or that Talbot's law is obeyed, assuming the inverse-square law to hold. Actually, the error due to any failure of L_2 to obey the inverse-square law is too small to be detected over the range of distances used. It might be noted here that when the transmission of thick specimens is measured, correction is made to the distance as read on the scale, this reading being larger than the true optical distance of L_2 from S_2."

Errors of observation are small because the final result is not a question of judgment, as in visual and photographic methods.

Errors to which all methods of spectrophotometry are liable, such as those resulting from stray radiant energy and inaccurate wave-length calibration, were carefully guarded against. Slits of 0·20 mm. (S_2 and S_3) have been usually used. This includes approximately from 1 $m\mu$ in the violet up to 4 $m\mu$ in the red. Errors due to these finite slit-widths are considered negligible except for very narrow bands.

§ (11) The Thermoelectric Method.—All that has been done in this connection is to adapt one of the well-known radiometric methods[3] to the measurement of spectral transmissive properties in the red and infra-red as far as 1350 $m\mu$. This limit was necessitated in this case because the thread of the spectrometer was cut no farther. The same spectrometer, glass prism, was used as already described in the previous section. A Coblentz linear thermopile and sensitive galvanometer were used. The apparatus has been arranged so that the thermopile and galvanometer may be substituted, in a moment's time, for the photoelectric cell P_2 (*Fig.* 5) and electrometer. L_1 is not used. Radiant energy from L_2 is focussed upon the slit S_2 and measurements made by moving the specimen in and out of the beam. The following points are of interest:

[1] Karrer and Tyndall, *Bur. Stand. Sci. Pap.* No. 389, 1920, p. 388. In this work the Brace spectrophotometer was used to measure the spectral transmission of the atmosphere.

[2] Ives, *Astrop. Jour.*, 1916, xliii. 24; Kunz, *Astrop. Jour.*, 1917, xlv. 69.

[3] See, for example, Coblentz, *Carnegie Inst. of Wash.*, Pub. No. 35, 1905.

(i.) The absorption of 1 cm. of water becomes so great at 1350 mμ that readings could not be made with any precision much beyond that point. The ordinary glass spectrometer is therefore sufficient for work upon aqueous solutions.

(ii.) Over this range and for a considerable way beyond, where large glass spectrometers and high-power incandescent lamps may be used, such an arrangement seems better than the usual infra-red spectrometer, because it is possible to work with very narrow slits. In this particular case slits of 0·20 mμ width were sufficiently large, that being equivalent to about 4 mμ in the red and about 18 mμ at 1000 mμ, less than 1/5 of the widths often used. This is important in the study of narrow absorption bands.

(iii.) Instead of a single lens it would be better to use two lenses, the light between them being in a parallel beam and the specimen inserted in this parallel beam. No errors because of the thickness of the specimen would then arise.

(iv.) The quartz-mercury lamp is excellent for calibration purposes in this range, it having strong lines at 1014 and 1129 mμ and weaker lines above 1300 mμ.

(v.) The method is of great aid to visual work between 650 and 800 mμ.

§ (12) The Photographic Method with the Hilger Sector Photometer and Auxiliary Apparatus.—A great advance in the ease and accuracy of photographic spectrophotometry has been made possible by the sector photometer, manufactured by Adam Hilger, Ltd. While this might be used in any spectral region where photography is applicable, its primary importance is in the ultra-violet. The measurement of the spectral transmissive properties of all kinds of organic and inorganic chemicals is of great importance to the chemist and the physicist, theoretically and practically, and an immense amount of work has been done along this line. The great outstanding fault, however, of most of this work in the ultra-violet is that it is merely qualitative. As noted by Hilger, "it must not be forgotten that not only the intensity but the actual position of the maximum of an absorption band is undetermined until quantitative measurements have been made, since any variations throughout the spectrum of intensity of the light source, or sensitiveness of the photographic plate, or of dispersion, may cause the apparent maximum of absorption to be in a position different from that of the actual maximum." While quantitative work is possible by the older photographic methods,[1] it is apt to be tedious and liable to serious errors which are entirely eliminated by the sector-photometer method.

[1] Résumé, Ewest, *Photog. Jour.*, 1914, liv. 90.

Auxiliary apparatus is necessary consisting of a quartz spectrograph and a source of ultra-violet radiant energy. The following descrip-

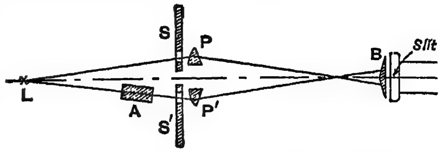


Fig. 6.—Hilger Sector Photometer Method.

tion of the method is taken from a paper published by Messrs. Hilger.

"Immediately in front of the slit and attached to it is a bi-prism, B, which receives the light from the source in the following manner. The light emanates from the source, L, and may reach the slit by two alternative paths. An upper beam passes through a rotating sector, S, the aperture of which can be varied. The beam then passes through the wedge prism, P, and, falling on the bi-prism, is deviated by the lower half of that prism to pass axially along the collimator of the spectrograph. The second beam traverses first the absorbing liquid under examination contained in a suitable cell, A, then through a rotating sector, S′, of fixed aperture, and a wedge prism, P′, similar to P but so arranged as to divert the light upwards instead of downwards. The beam is then diverted by the upper half of the bi-prism, B, and passes axially along the collimator of the spectrograph like the first. We thus have the spectrograph fed by two beams, the one capable of being varied in intensity at will by varying the aperture of the rotating sector, the other subject to the absorption of a known thickness of the liquid under examination. A series of photographs is taken with the sector S set to different apertures. If we consider one of these photographs we shall see that it consists of a pair of spectrum photographs in close juxtaposition, one of which is of reduced density throughout its whole length, the other—that, namely, which has passed through the material under test—being more dense than the first in certain parts and less so in others, there being certain wave-lengths where the density of the two is equal.

"It is found that the wave-length at which the densities of the two spectra are equal is, within wide limits, independent of the exposure, of the intensity of illumination, and of the speed of rotation of the sectors.

"Furthermore, the photographs are taken simultaneously, and there are, therefore, no errors arising from fluctuations in the light source.

"Let the fractions of whole revolutions

during which the variable and fixed aperture sectors allow light to pass be respectively t and t', and let the intensities of light of wave-length λ reaching the slit by the upper and lower beams in *Fig.* 6 be respectively I and I′.

"Then if at wave-length λ the density is exactly the same in both spectra, we have

$$\frac{I}{I'}=f(t', t).\text{"}$$

The makers of the instrument assume the Schwarzchild[1] relation, $I/I'=(t'/t)^n$ for uninterrupted exposures, to hold in this case, n being different from unity, and supply a calibration from which I'/I may be obtained from the known values of t/t'. It has been found,[2] however, that "the photographic plate, the use of which in photometric measurements is usually considered questionable, integrates intermittent exposures in such a way that the comparison of two intensities can be made directly in terms of sector openings, provided that the time from the beginning to the end of the exposures is the same, and that the two exposures produce equal blackening of the plate"; *i.e.* the constant n for the sector photometer is unity within the limits of photometric accuracy. This relation has been further verified by comparison with other methods,[3] and extensively by comparison with the König-Martens visual method and the photoelectric null method at the Bureau of Standards. These data have been published in the *Bureau of Standards Paper*, No. 440, on the spectral transmissive properties of the seven food dyes.

The sector photometer at the Bureau of Standards has both sectors variable, so that, for example, the cells containing solution and solvent may be interchanged for check readings. The quartz spectrograph is a Fuess instrument (slit-width = 0·10 mm.) with a dispersion of nearly 14 cm. between 500 and 250 mm. The source of radiant energy is a high-voltage (Tesla coil) spark under water, which gives a continuous spectrum from the visible as far as the quartz system will transmit. The vertical spark is more satisfactory than the horizontal, for it keeps the two spectra in contact throughout their whole length. Two pieces of magnesium carbonate diffusely illuminated as in the König-Martens auxiliary apparatus, but by the spark, would be still better and would enable measurements of diffuse spectral reflection to be made as well, provided sufficient intensity could be obtained.

Errors in locating the wave-length or frequency of equal density have been made very small by means of a special comparator, described in Appendix to *Bur. Stand. Tech. Pap.* No. 148, which permits of readings being made without marking the negative. Therefore any number of readings may be made and an average taken. The principal source of error at present is thought to be due to occasional non-uniformity of the spark. Displacement of the electrodes, and therefore of the spark as a whole, from the proper position may be quite accurately controlled. Interchange of solution and solvent will test this, as well as exposure with both cells containing solvent.

[1] *Astrop. Jour.*, 1900, xi. 89.

[2] Howe, *Phys. Rev.* (2), 1916, viii. 674.

[3] *Trans. Opt. Soc.*, 1917, xviii. 36; *Phys. Rev.* (2), 1917, x. 707; *Bur. Stand. Tech. Pap.* No. 119, 1919, and No. 148, 1920.

§ (13) General Comparison and Relative Accuracy of the Different Methods.—The agreements and lack of agreements found when the spectral transmissions of a large number of glasses were first obtained by different methods are well illustrated in *Bur. Stand. Tech. Pap.* No. 119. In this investigation four methods were used: the König-Martens and the Lummer-Brodhun spectrophotometers in the visible, and also the Martens photometer with selective ray filters and monochromatic light, and the Hilger sector photometer method in the blue, violet, and ultra-violet. Each apparatus was considered to be in reliable working condition, but the comparison of results showed occasional very serious discrepancies, especially in the blue and in the red beyond 650 mμ. Much of this was due to insufficient precautions against stray light.

Since that time the four methods already described in detail—viz. the visual, the photoelectric, the thermoelectric, and the photographic—have been developed or improved; and values of spectral transmissive characteristics may now be obtained from about 230 to 1350 mμ with no "weak" regions and no serious discrepancies. In the paper on the food dyes noted above the observed values of transmittancies[4] by all four methods are plotted for all the concentrations, thicknesses, and temperatures studied. As a result of this investigation the following statements may be made:

(i.) Errors of method are believed to have been eliminated in the visual method with the König-Martens spectrophotometer when homogeneous light is used.

(ii.) There is very close agreement between the visual and photoelectric methods. This is true not only between 500 and 550 mμ, where both methods are very reliable, but throughout nearly the whole range of either.

(iii.) In comparing the photographic determinations with the visual or photoelectric between 390 and 500 mμ, it has been found that the agreement is usually better the shorter the wave-length, the less the transmittancy, and the "steeper" the transmittancy curve. From the nature of the photographic method it is difficult to determine exactly the value of maximum or minimum

[4] *Transmittancy*: the ratio of the transmission of a cell containing a substance in solution to that of the same or an equivalent cell filled with the solvent only.

transmittancy. But the agreement in the blue and violet indicates that in the ultra-violet, where the dispersion is greater, the method is very reliable, especially for the determination of the frequency and magnitude of *absorption* bands, the main purpose of most investigations in this region.

(iv.) Agreements between the visual and thermoelectric methods between 600 and 700 mμ indicate that no serious errors are present.

(v.) On the König-Martens spectrophotometer with homogeneous light, values of k, the specific transmissive index,[1] may, as a rule, be obtained accurate to within ± 2 per cent, provided that suitable concentrations and thicknesses are used.[2]

(vi.) Values of k at any wave-length between about 400 and 680 mμ visually or photoelectrically (*i.e.* both methods being used) with a heterogeneous source may be obtained with slightly less accuracy than this.

(vii.) Definite maximum values of k, *i.e.* the absorption bands, if not too narrow, may be obtained in the visible accurate to within ± 3 per cent; but in the ultra-violet, because of certain characteristics of the method, the uncertainty may often be greater than this.

(viii.) In the visible and ultra-violet the wave-lengths of definite maximum values of k are accurate within ± 1 mμ; in the infra-red the uncertainty becomes greater, but is considered less than ± 5 mμ even at 1350 mμ.

The visual method is the only one so far developed by which measurements of diffuse spectral reflection may be made under approximately ideal conditions, and there is no reason to doubt that values of reflection are obtained with as high accuracy as is possible in transmissive work. As already noted, the photoelectric method gave values of reflection considered of perhaps half the accuracy of transmissive measurements, for approximately diffusing surfaces. Examples of the close agreement between the visual and photoelectric methods for such surfaces are given in *Bur. Stand. Tech. Pap.* No. 167. Approximate measurements on some of the same samples by the photographic sector photometer method with a quartz-mercury lamp as source indicate the reliability of the method if proper irradiation is used.

K. S. G.

SPECTRO-POLARIMETER: an instrument for measuring the rotation for different wave-lengths of the plane of polarisation of light by various substances. See "Polarimetry," § (14).

[1] *Specific transmissive index*, k: Let T = transmittancy, c = concentration, b = thickness; then $k = -\log_{10} \sqrt[cb]{T}$.

[2] The photometric precision of this instrument, as tested by means of accurately calibrated rotating sectors, is better than 1 per cent.

SPECTROSCOPES AND REFRACTOMETERS

I. Theory

§ (1).—When light falls on a prismatic piece of a transparent substance it is deviated by an amount depending on the angle of the prism, and on the property of the substance known as its refractive index, which varies with the wave-length of the light. Light of different wave-lengths is therefore deviated to different extents; and if the incident beam contains constituents of more than one wave-length, each constituent beam emerges from the prism in a different direction.

The deviation of a beam of light by a prism depends, therefore, on two inter-related physical quantities, the refractive index of the prism and the wave-length of the light; and it may be measured with a view to determine unknown wave-lengths, if the properties of the prism are known (or eliminated), or to determine the refractive properties of the prism for known wave-lengths. The instrumental requirements for precise determinations vary somewhat, depending on which of these purposes is in view; and the spectroscopist, who is primarily interested in wave-length measurements, would rarely have much use for the instrument best suited to the requirements of the refractometrist, who is primarily interested in the refractive indices of the prism.

Since the determination of refractive indices may be said to be the basic measurement in applied optics, inasmuch as the design of any optical instrument depends for its realisation on a precise knowledge of the refractive properties of the materials used in its construction, we shall devote some consideration to the methods and instruments best adapted to this measurement; and in the first place we shall investigate generally the conditions which govern the refraction of light by a "prism" with a view to determining the comparative accuracy of different methods and the actual accuracy likely to be obtained with a given instrumental equipment.

§ (2) Passage of Light through a Prism.—All the methods suitable for precise refractometry involve the determination of the angular deviation, with respect to some fixed direction, of a beam of light which has been refracted by a system of plane surfaces, usually two in number. Separate formulae are readily obtained for the various methods; but it is more useful, in order to correlate such methods and compare their advantages and disadvantages, to treat them as special cases of a quite general relationship.

Let AB, *Fig.* 1, be a ray of monochromatic light incident on the first of two plane surfaces

which separate three regions in which the refractive indices for the wave-length of the light employed are μ_1, μ_2, and μ_3. We shall assume that the incident ray and also the normals to the surfaces lie in the plane of the paper.

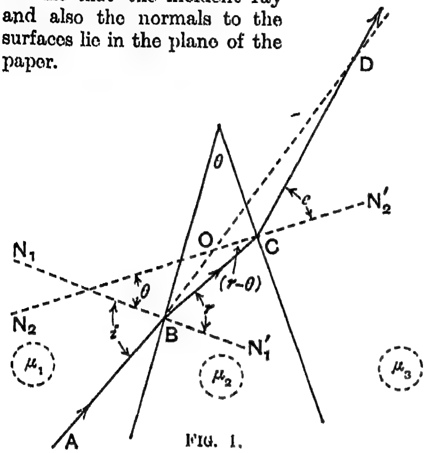


FIG. 1.

It is necessary to employ a definite convention with respect to the signs of angles in order to prevent confusion, and, in certain cases, ambiguity. *All angles will be considered positive when measured in a counter-clockwise direction from the appropriate reference directions.* For angles of incidence and refraction the reference directions are the normals to the respective surfaces; while, in considering the angle between two surfaces, the normal to the surface first encountered by the ray is the reference direction for the normal to the second surface. As regards the deviation of a ray, the initial direction of the ray is the reference line for the final direction.

Fig. 1 has been drawn so that all the angles with which we are concerned are positive on the above convention. The formulae obtained from it will of course differ in the signs of certain terms from those deduced in the text-books, in which the magnitude of an angle is alone taken into account.

Let the angle of incidence $ABN_1 = i$. The refracted ray BC in the second medium makes an angle $N_1'BC$, $= r$, with the normal at B, and finally enters the third medium, making an angle of emergence $N_2'CD$, $= e$, with the normal to the second surface at C. If the angle between the surfaces is θ, it follows from the geometry of the figure that $B\hat{C}N_2$, the angle of incidence at the second surface, $= r - \theta$, and that angle $C\hat{D}O = \theta + e - i$. But $C\hat{D}O$ is the angle by which the ray has been deviated by the two refractions, $= \delta$ say. Therefore the deviation, δ, $= \theta + e - i$.

$$\sin^2\theta = \{\sin r \cos(r-\theta) - \cos r \sin(r-\theta)\}^2$$
$$= \sin^2 r + \sin^2(r-\theta) - 2\sin r \sin(r-\theta)\cos\theta.$$

From the law of refraction

$$\mu_1 \sin i = \mu_2 \sin r\ ;\ \mu_2 \sin(r-\theta) = \mu_3 \sin e\ ;$$

therefore

$$\mu_2^2 \sin^2\theta = \mu_1^2 \sin^2 i + \mu_3^2 \sin^2 e - 2\mu_1\mu_3 \sin i \sin e \cos\theta. \quad (1)$$

Equation (1) is the starting-point from which all special cases may be deduced.

(i.) *Parallel Surfaces.*—If $\theta = 0$, equation (1) reduces to

$$\mu_3 \sin e = \mu_1 \sin i.$$

But this is the same relationship between i and e that would apply if the first and third media were separated by a single surface. The parallel plate of index μ_2 is therefore without effect on the final direction of the beam, which depends simply on the properties of the first and last media. If $\mu_1 = \mu_3$ the ray will be undeviated by passing through the plate.

(ii.) *Prism in Air.*—In this case $\mu_1 = \mu_3 = 1$ (approx.), and equation (1) becomes

$$\mu_2^2 \sin^2\theta = \sin^2 e + \sin^2 i - 2\sin e \sin i \cos\theta. \quad (2)$$

There are three special cases of practical importance, viz. $e = 0$; $i = 90°$; and the case in which the deviation is a minimum.

(*a*) When $e = 0$ the ray strikes the second surface of the prism normally. Equation (2) reduces in this case to

$$\mu_2 = \frac{\sin i}{\sin\theta} \quad . \quad . \quad . \quad (3)$$

(*b*) When $i = 90°$ the incidence is "grazing," the ray being refracted at the so-called critical angle.

Equation (2) reduces in this case to

$$\mu_2^2 = 1 + \left(\frac{\cos\theta - \sin e}{\sin\theta}\right)^2. \quad . \quad (4)$$

(*c*) When the deviation is a minimum the angles of incidence and emergence are equal. This is shown in all text-books and it is needless to prove it here. When this condition is fulfilled $i = -e$ and equation (2) becomes

$$\mu_2^2 \sin^2\theta = 2\sin^2 i(1 + \cos\theta),$$

or $4\mu_2^2 \sin^2(\theta/2)\cos^2(\theta/2) = 2\sin^2 i \,.\, 2\cos^2(\theta/2)$,

whence $$\mu_2 = \frac{\sin i}{\sin(\theta/2)}.$$

The deviation, δ, $= \theta + e - i$, $= \theta - 2i$, or $i = \frac{1}{2}(\theta - \delta)$; therefore

$$\mu_2 = \frac{\sin\frac{1}{2}(\theta-\delta)}{\sin\theta/2} \quad . \quad . \quad . \quad (5)$$

The reader may have observed that the sign of $\sin e$ in equation (4) and of δ in equation (5)

are different from the usually given forms. In the latter case, δ is invariably negative on the sign convention adopted in this treatment; so if $\bar{\delta}$ denotes the numerical value of the measured deviation, $\delta = -\bar{\delta}$ and

$$\mu_2 = \frac{\sin \frac{1}{2}(\theta + \bar{\delta})}{\sin \theta/2}, \quad . \quad . \quad (5a)$$

which is the usual form. In the case of equation (4) the formula is usually deduced in the text-books for a case in which e is negative, in which case, if the numerical value of $\sin e$ is denoted by $\sin \bar{e}$, the usual $+$ sign is obtained in the formula. However, in this method e is by no means always $-$ve. If θ or μ_2 is small, the emergent ray may lie between the normal and the refracting edge of the prism, in which case e is positive and the negative sign in equation (4) must be used. It is easier therefore, as in other optical problems, to remember the generalised forms of these equations, inserting the appropriate signs in practical cases, rather than to remember each case separately.

(iii.) *Prism with one Surface in Contact with another Refractive Substance.*—Suppose that in contact with the first surface of the prism there is another refractive substance. In this case $\mu_1 \neq 1$, and equation (1) becomes

$$\mu_2^2 \sin^2 \theta = \sin^2 e + \mu_1^2 \sin^2 i - 2\mu_1 \sin i \sin e \cos \theta.$$

The only case of practical importance is that of grazing incidence, or $i = 90^\circ$.

In this case the equation reduces to

$$\mu_1 = \pm \sin \theta \sqrt{\mu_2^2 - \sin^2 e} + \sin e \cos \theta.$$

The index μ_1 is positive and greater than unity, whereas the product $\sin e \cos \theta$ is necessarily less than unity; so that the positive sign before the square root must always be taken. Thus

$$\mu_1 = \sin \theta \sqrt{\mu_2^2 - \sin^2 e} + \sin e \cos \theta. \quad (6)$$

If μ_2 and θ are known, μ_1 can be determined from equation (6), the only measurement required being the angle of emergence e which the ray makes with the normal to the second surface of the prism.

§ (3) Angle of Incidence for a given Deviation.—In the preceding section the relations involved in the principal methods of refractometry have been deduced. In each case some particular value of i, e, or δ was chosen, the choice being dictated by the convenience of realising these particular conditions in practice.

It is sometimes desired to produce a given deviation by a given prism (in air), and an angle of incidence has to be chosen to give this deviation.

The general equation, (2), is

$$\begin{aligned}\mu_2^2 \sin^2 \theta &= \sin^2 e + \sin^2 i - 2 \sin e \sin i \cos \theta \\ &= \tfrac{1}{2}\{2 - (\cos 2e + \cos 2i)\} \\ &\qquad - \cos \theta \{\cos (e - i) - \cos (e + i)\} \\ &= \tfrac{1}{2}[2 - 2\{\cos (e + i) \cos (e - i)\}] \\ &\qquad - \cos \theta \cos (e - i) + \cos \theta \cos (e + i) \\ &= 1 - \cos (e + i)\{\cos (e - i) - \cos \theta\} \\ &\qquad - \cos \theta \cos (e - i).\end{aligned}$$

But $\quad e + i = \delta + 2i - \theta$, and $e - i = \delta - \theta$;

$$\therefore \mu_2^2 \sin^2 \theta - 1 + \cos \theta \cos (\delta - \theta) = -\cos (\delta - \theta + 2i)\{\cos (\delta - \theta) - \cos \theta\}$$

or

$$\cos (\delta - \theta + 2i) = \frac{1 - \mu_2^2 \sin^2 \theta - \cos \theta \cos (\delta - \theta)}{\cos (\delta - \theta) - \cos \theta}.$$

Since δ is always $-$ve, provided $\mu_2 > 1$, we may put $\delta = -\bar{\delta}$ where $\bar{\delta}$ is the numerical value of the deviation, and

$$\cos (2i - \theta - \bar{\delta}) = \frac{1 - \mu_2^2 \sin^2 \theta - \cos \theta \cos (\theta + \bar{\delta})}{\cos (\theta + \bar{\delta}) - \cos \theta}.$$

from which, since θ and μ_2 are supposed known, the value of i corresponding to a specified value of δ can be calculated.

§ (4) Sensitivity of Methods.—The equations of § (2), relating to various methods of refractometry employed in practice, tell us nothing about the relative advantages of the methods, and give no guidance as to which should be used in any particular circumstances. It is necessary to discuss the sensitivity of the methods, that is, the change in the measured quantity per unit change of refractive index, and also the extent to which the results are affected by errors in the values of the auxiliary constants, such, for instance, as the angle of the prism, or, in the case of the method to which equation (6) refers, the refractive index μ_2 of the auxiliary prism.

Taking first the most familiar method, the quantity measured is δ, the minimum deviation. The sensitivity S may be measured by the number of seconds which δ varies when the refractive index of the prism changes by one unit in the fifth decimal place. Thus

$$S = \frac{\partial \delta}{\partial \mu_2} \times \frac{57 \cdot 3 \times 60 \times 60}{10^5} = 2 \cdot 06 \frac{\partial \delta}{\partial \mu_2}.$$

From equation (5) above,

$$\frac{\partial \mu_2}{\partial \delta} = -\frac{1}{2} \frac{\cos \frac{1}{2}(\theta - \delta)}{\sin \theta/2}.$$

Whence $$S = \frac{-4 \cdot 12 \sin \theta/2}{\cos \frac{1}{2}(\theta - \delta)}. \quad . \quad . \quad (7)$$

Thus the sensitivity is a function of the deviation and of the angle of the prism. We can calculate its value for any values of μ_2 and θ, first finding the appropriate values of δ from equation (5).

For a given material (*i.e.* μ_2 constant) it is interesting to observe how the sensitivity varies for different prism angles. There is clearly a major limit to the permissible angle for any value of μ_2, which is reached when θ is such that both the angles of incidence and emergence are 90°. If θ is greater than this there is no emergent ray, the light being totally reflected at the second surface for all possible angles of incidence on the first.

If $i = -e = 90°$, $\delta = \theta + e - i = \theta - 180°$, and equation (5) reduces to

$$\mu_2 = \frac{1}{\sin \theta/2}.$$

The largest refractive index of which measurement can be made with a prism of given angle, or the largest angle which a prism of given index may have and still transmit light, is given by the above relationship.

If $\mu_2 = 1{\cdot}6$ the largest possible prism angle is $77\frac{1}{3}°$, while for the usual 60° prism the largest possible index is 2.

In *Fig.* 2 the sensitivity of prisms of refractive index 1·6 and 1·75 is plotted for different prism angles. The sensitivity increases as θ increases, the rate of increase being very rapid as θ approaches its greatest possible value, being infinite when that value is reached. The curve shows that for $\mu_2 = 1{\cdot}6$, which is about the average of the values encountered in practical work of high precision, an angle of 75° would give more than five times the sensitivity of the usual 60° prism.

Sensitivity is, however, only one factor in the accuracy of the measurements; and we shall see that there are other factors which prevent the utilisation of the large sensitivity which might be obtained by suitable choice of prism angle.

Leaving for the moment the case of minimum deviation, let us consider the method in which the incidence is grazing, and for which we found (equation (4))

$$\mu_2^2 = 1 + \left(\frac{\cos \theta - \sin e}{\sin \theta}\right)^2.$$

In this case the measured angle is e, and the sensitivity, in the same units as adopted for the last case, will be $S = 2{\cdot}06\,(\partial e/\partial \mu_2)$.

$$2\mu_2 \frac{\partial \mu_2}{\partial e} = -2\left(\frac{\cos \theta - \sin e}{\sin \theta}\right)\frac{\cos e}{\sin \theta},$$

or $$S = -2{\cdot}06\mu_2 \frac{\sin^2 \theta}{\cos e}\Big/(\cos \theta - \sin e). \qquad (8)$$

The quantity in brackets is always positive provided $\mu_2 < 1$, since, if μ_2 is unity, the ray is undeviated and emerges parallel to the 1st face, in which case $e = 90° - \theta$ and $\sin e = \cos \theta$.

This is the largest positive value e can have, so that if $\mu_2 > 1$, $\sin e < \cos \theta$. Thus $\partial e/\partial \mu_2$ is always −ve and the ray is always deviated farther away from the refracting edge for an increase in the refractive index.

Clearly the maximum value of θ for a given μ_2, or *vice versa*, is the same as in the minimum deviation method, since the case in which *both* i and $-e = 90°$ may be regarded indifferently as pertaining to either method.

As regards the sensitivity obtained with different angles for a given μ_2, the general features of the curve would be similar to *Fig.* 2,

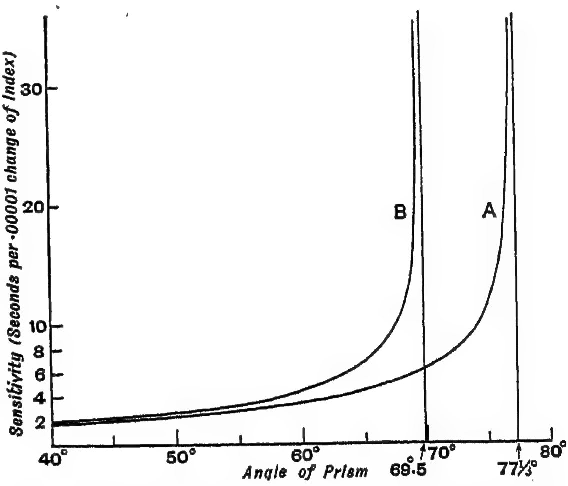


FIG. 2.—Sensitivity of Minimum Deviation Method for different Prism Angles.
A, $\mu = 1{\cdot}60$. B, $\mu = 1{\cdot}75$.

for it is easy to see from the form of equation (8) that as $\cos e$ approaches zero, that is, as the angle of the prism approaches the value for which the emergent ray is grazing, $\partial e/\partial \mu_2$ increases more and more rapidly, becoming infinity when the limiting value of the angle is reached. In this case as in the last, however, practical considerations prevent the highest sensitivity being fully utilised and, as we shall see, the balance of advantage for most ordinary purposes lies with the 60° prism.

If then we take the 60° prism as basis, it is useful to compare the sensitivities of these two methods throughout the range of refractive index usually encountered in practice.

These have been plotted in *Fig.* 3. The curves tell us the accuracy with which the direction of the emergent beam has to be determined in each method in order to have an error not exceeding 0·00001 in the refractive

index. The precision required is greatest for low indices. For indices below 1·5, an error of 3 seconds in the direction of the emergent ray will introduce an error of 0·00001 in the result with either method. At higher indices there is a slight improvement, particularly with the minimum deviation method. The region of high sensitivity (1·8 to 2·0) is not of great importance, as most substances employed in optical work have indices between 1·4 and 1·8, the great majority lying between

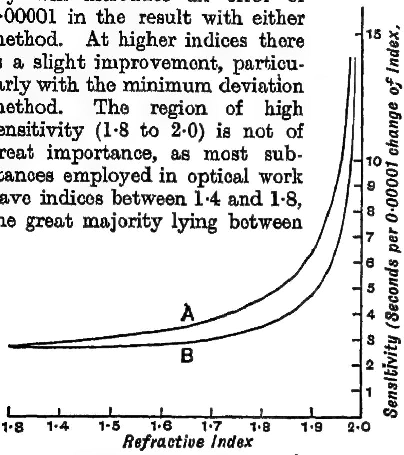


FIG. 3.—Sensitivities of Refractive Index Measurements for different Indices.

A, Minimum Deviation Method.
B, Grazing Incidence Method.

1·5 and 1·75. In the majority of cases, then, the sensitivity for the grazing incidence method is 3 seconds per 0·00001 change of index and for the minimum deviation method from 3 to 4 seconds.

The third method which it is of importance to discuss is that due to Wollaston,[1] in which the object is not the determination of the index of the prism but of a substance in contact with its first face, the prism merely filling an auxiliary rôle. Several commercial refractometers due to different inventors are made on this principle. The prism we may term the standard prism or standard block.

Equation (6) refers to this method, μ_1 being the index to be determined, μ_2 the index, for the wave-length in question, of the standard block, and θ its angle. The quantity measured is e, the angle of emergence from the second face of the block.

Differentiating equation (6) under the conditions that θ and μ_2 are constant, we obtain

$$\frac{\partial\mu_1}{\partial e}=\frac{1}{2}\frac{\sin\theta}{\sqrt{\mu_2^2-\sin^2 e}}(-2\sin e\cos e)+\cos e\cos\theta$$

$$=-\frac{\cos e}{\sqrt{\mu_2^2-\sin^2 e}}\{\cos\theta\sqrt{\mu_2^2-\sin^2 e}-\sin e\sin\theta\},$$

which, after transformation and substitution from equation (6), becomes

$$\frac{\cos e}{\sqrt{\mu_2^2-\sin^2 e}}\sqrt{\mu_2^2-\mu_1^2}.$$

Thus the sensitivity for measurements of μ_1, which we may denote by $S\mu_1$,

$$=2\cdot06\frac{\partial e}{\partial\mu_1}=\frac{2\cdot06\sqrt{\mu_2^2-\sin^2 e}}{\cos e\sqrt{\mu_2^2-\mu_1^2}}. \quad (9)$$

This quantity can be calculated for different values of μ_1, μ_2, and θ, the appropriate values of e being obtained from equation (6), which, for this purpose, can be reduced to a rather more convenient form, viz.

$$\sin e=\mu_1\cos\theta-\sin\theta\sqrt{\mu_2^2-\mu_1^2}. \quad (10)$$

In *Fig.* 4 a series of curves are drawn showing, for the case in which $\mu_2=1\cdot75$ (a usual value for this method), how the sensitivity, $S\mu_1$, varies with the index under test (μ_1) for prisms of different angles. The sensitivity is in all cases high for indices near that of the block, diminishing to a minimum value, which is lower, and occurs at a lower value of μ_1, the smaller θ becomes. Thus with a 90° block, the minimum sensitivity occurs when $\mu_1=1\cdot57$ approx., and is about 6 seconds per 0·00001 variation of μ_1; for $\theta=80°$ the minimum is at $\mu_1=1\cdot5$, and is under 5 seconds; for $\theta=60°$ the minimum is not reached until μ_1 is in the neighbourhood of 1·2, and is just over 3 seconds, while for $\theta=45°$ the minimum is never reached.

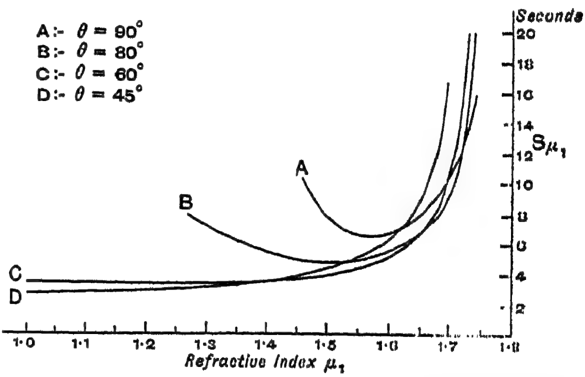


FIG. 4.—Sensitivity of Refractive Index Measurements by Wollaston Method.

§ (5) EFFECTS OF ERRORS IN AUXILIARY CONSTANTS.—In all the formulae which we have considered the angle, θ, of the prism occurs. This quantity has, in general, to be experimentally determined, and is therefore subject to possible error. The assumption of an erroneous value for θ in any of the

[1] *Phil. Trans.*, 1802, p. 365.

formulae for refractive index will lead to a corresponding error in the result. This error is clearly proportional to the partial derivatives with respect to θ of the expressions for μ.

For the minimum deviation method we obtain, by partial differentiation of equation (5),

$$\frac{\partial\mu_2}{\partial\theta}=\tfrac{1}{2}\frac{\cos(\theta-\delta/2)}{\sin\theta/2}-\tfrac{1}{2}\frac{\sin(\theta-\delta/2)}{\sin^2\theta/2}\cos\theta/2$$

$$=\tfrac{1}{2}\frac{\sin\delta/2}{\sin^2\theta/2}.$$

It is usually more convenient to express the result as the inverse of this, *i.e* the error in angle which will produce unit error in the calculated refractive index. We may call this the *Tolerance in prism angle* and denote it by T_θ, the factor 2·06 being introduced as before to give the result in seconds of arc per 0·00001 error in index.

Thus $$T_\theta=4{\cdot}12\frac{\sin^2\theta/2}{\sin\delta/2}. \qquad (11)$$

For the grazing incidence method by differentiation with respect to θ, of equation (4) we obtain

$$2\mu_2\frac{\partial\mu_2}{\partial\theta}=2\left(\frac{\cos\theta-\sin e}{\sin\theta}\right)\left\{-\frac{\sin\theta}{\sin\theta}-\left(\frac{\cos\theta-\sin e}{\sin^2\theta}\right)\cos\theta\right\}$$

$$=-2\left(\frac{\cos\theta-\sin e}{\sin\theta}\right)\left\{1+\cos\theta\left(\frac{\cos\theta-\sin e}{\sin^2\theta}\right)\right\}$$

$$=-2\{\sqrt{\mu_2^2-1}+(\mu_2^2-1)\cot\theta\},$$

$$T_\theta=2{\cdot}06\frac{\partial\theta}{\partial\mu_2}=\frac{-2{\cdot}06\,\mu_2}{\sqrt{\mu_2^2-1}+(\mu_2^2-1)\cot\theta}. \qquad (12)$$

In *Fig.* 5, T_θ is plotted for both of these methods for different prism angles, the refractive index being taken as 1·6. We see

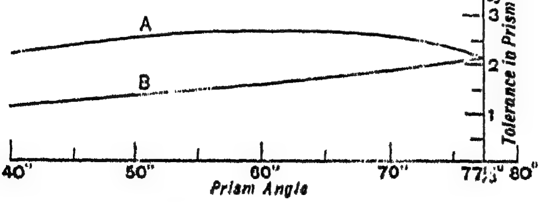


FIG. 5.—Tolerance in Angle of Prisms used for Refractive Index Determinations.
A, Minimum Deviation } μ=1·6.
B, Grazing Incidence }

from the figure that the effect of an error in the angle of the prism is much greater for grazing incidence than for minimum deviation, except near the limiting prism angle, in which

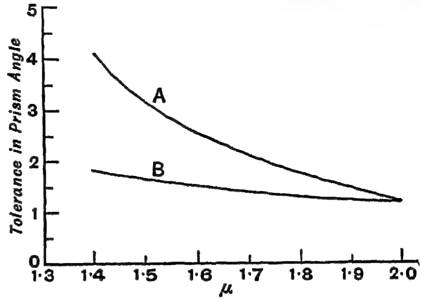


FIG. 6.—Tolerance in Angle of Prism used for Refractive Index Measurements.
A, Minimum Deviation.
B, Grazing Incidence.

case the two methods tend to become identical. In *Fig.* 6 is plotted the tolerance with a 60° prism, for different values of the index.

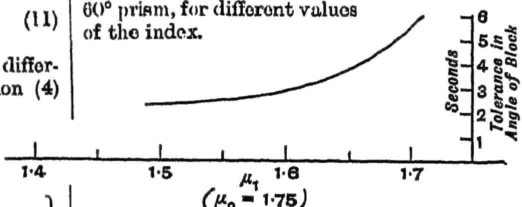


FIG. 7.—Tolerance of Angle of Block in Wollaston Method.

In the case of the Wollaston method we have, from equation (6),

$$\frac{\partial\mu_1}{\partial\theta}=\cos\theta\sqrt{\mu_2^2-\sin^2 e}-\sin e\sin\theta,$$

which reduces to $\sqrt{\mu_2^2-\mu_1^2}$.

Thus $$T_\theta=2{\cdot}06\frac{\partial\theta}{\partial\mu_1},=\frac{2{\cdot}06}{\sqrt{\mu_2^2-\mu_1^2}}. \qquad (13)$$

In this case, therefore, the tolerance in angle is independent of the angle of the prism, depending only on the values of μ_1 and μ_2. Assuming $\mu_2=1{\cdot}75$, *Fig.* 7 shows the value of T_θ for different values of the index under test. We see from it that the effect of an error in θ is less the higher the index under test. For substances of index about 1·5 an error of two seconds in the angle of the block will introduce an error of 0·00001 in the value of the index.

The accuracy of this method also depends on the accuracy with which μ_2, the index of the prism, is known. The error in μ_1 due to an error in the assumed value of μ_2 will be proportional to the partial derivative of the

The use of quartz jaws was suggested by the late Sir W. Crookes.[1] These are bevelled to a very sharp edge in the same way as the ordinary metal jaws. The edges therefore form prisms, and light which passes through the quartz is refracted towards the sides of the collimator tube, by which it is absorbed. Only the light passing between the edges reaches the lens of the collimator, just as in the case of opaque jaws. The advantage of quartz jaws is that they can be polished to a more accurate edge than metal. The disadvantage is that the edges are so extremely fragile. The jaws should therefore be mounted in such a way that the moving mechanism drags them forward by friction, so that there is no possibility of jamming them together in the event of the slit being accidentally closed. It is surprising that this method has not been generally adopted even for metal jaws, as it would greatly minimise the risk of damage to a valuable slit.

The slit is usually attached to a short tube which slides into the end of the collimator tube, and may be adjusted, preferably by rack and pinion, for focussing the collimator.

The body tube of the collimator should be lined with some corrugated material or fitted with a series of diaphragms to prevent light which falls on the side of the tube from being reflected as stray light through the lens.

The lens of the collimator should be a telescope objective of the finest quality. It should be free from all but the merest trace of spherical aberration. The lenses employed in spectrometers are almost invariably cemented doublets. There is therefore always considerable residual chromatic aberration, which is usually particularly manifest at the blue end of the spectrum. By the use of triple lenses the greater part of this could be obviated, and a very serious nuisance to the spectroscopist removed.

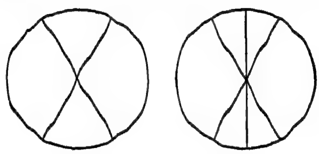
(ii.) *The Telescope.*—The requirements of this important part of the spectrometer have not, unfortunately, received the attention from makers which ought to have been devoted to them. As in the case of the collimator, telescopes are frequently fitted with object glasses made to an ordinary formula suitable for telescopes of surveying instruments, etc. The type of colour correction suitable for such instruments is entirely unsuited for spectroscopic work, in which the rays at the extreme ends of the spectrum are just as important as those in the more central regions. Here, as in the case of the collimator, it would be possible to get a much more satisfactory result from a triple lens than from the usual doublet.

As the telescope may sometimes have to be used with a micrometer eyepiece, the lens ought to be designed to give the best definition possible over a field of at least 2° on either side of the axis. Further, the field of view should be flat, otherwise the measurements with the micrometer eyepiece will be affected by distortion.

The power of the eyepiece employed may vary over a fairly wide range without appreciably affecting its usefulness; but a convenient power is one which gives a magnification of the telescope as a whole of 18 to 20 diameters for each inch of effective aperture.

An important feature of a spectroscopic telescope, and one which greatly influences the accuracy of the results obtainable with the instrument, is the cross-lines.

Considerable experimenting with settings of different types has been done by various workers, but there is very little doubt that the most suitable arrangement for general spectroscopic work is that in which the cross-lines are mounted diagonally as shown in *Fig.* 9 (*a*). The angle between the lines varies in different instruments from about 20° to 90°. Where possible the writer always mounts them at 60°. Such lines can be set with great accuracy on a vertical line as in *Fig.* 9 (*b*). As regards the thickness of the lines, there is no great advantage in having them excessively fine. Within reasonable limits, the thicker the lines the less fatiguing it is to keep them



(*a*) (*b*)

FIG. 9.

in sharp focus. With an eyepiece of about 1½ cm. focal length a convenient diameter of fibre to employ is from 4 to 8 μ.

In the case of a micrometer eyepiece, the best type is that in which the ocular can be moved as well as the cross-lines. If the ocular is fixed only small movements of the cross-lines are possible, since the definition falls off rapidly towards the edge of the field. But if the ocular can also be translated, so that the cross-lines always appear in the centre of the field of view, good definition is obtained over a much larger range.

(iii.) *The Axis and Scale.*—The considerations of §§ (4) and (5) show us that in order to measure refractive indices correctly to the fifth decimal place it is necessary that we should be able to determine angles to within a little over one second of arc. This requires most accurately turned centres and accurately divided scales; but even when the instrument-maker has done his very best, there is much to be done by the experimenter in discovering the peculiarities

[1] *Chem. News*, 1895, lxxi. 175.

right-hand side of equation (6) with respect to μ_2, θ and e being regarded as constant.

$$\frac{\partial\mu_1}{\partial\mu_2}=\frac{1}{2}\frac{\sin\theta}{\sqrt{\mu_2^2-\sin^2 e}}\cdot 2\mu_2=\frac{\mu_2\sin\theta}{\sqrt{\mu_2^2-\sin^2 e}}. \quad (14)$$

In the particular case when $\theta=90°$, this reduces to μ_2/μ_1. In *Fig.* 8, $\partial\mu_1/\partial\mu_2$ is plotted for different prism angles for two values of μ_1.

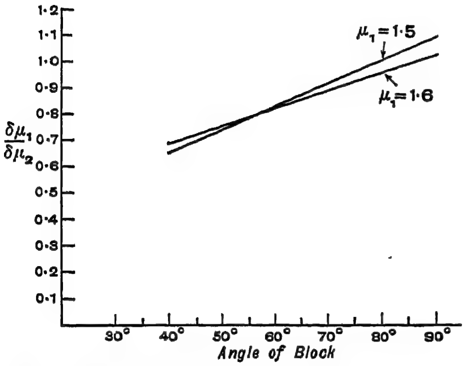


FIG. 8.—Wollaston Method. Effect of Errors in Index of Standard Block.

It will be noticed that the smaller the value of θ the smaller is the effect of an error in μ_2.

II. The Spectrometer

§ (6).—The spectrometer in its simplest elements consists of a collimator, an adjustable table on which a prism or other dispersing apparatus may be placed, a telescope which rotates about an axis concentric with the table, and a circular scale on which the angular position of the rotating telescope may be determined. It is necessary to be able to rotate the table itself about the same axis as the telescope rotates; and, except in the simplest instruments, means are provided for reading the angular position of the table, either on the same circle as used for the telescope or on a separate circle.

The former arrangement is the more generally useful, since it enables rotations of the table to be determined with the same accuracy as rotations of the telescope. Where a separate circle is provided for the table it is rarely of the same accuracy as the other.

(i.) *The Collimator.*—This consists simply of a narrow slit at the focus of a lens. The rays from each point of the slit are therefore rendered parallel by passage through the lens. The necessity for collimation is due to the fact that unless the slit were virtually at infinity, rays from any point of it would meet different parts of the prism at different angles of incidence and would therefore be deviated to different extents. If the slit is at infinity, all the rays from any point of it which reach the prism are deviated equally and can be recombined at a single point in the focal plane of the telescope. A sharp image of the slit is the result. The slit is usually formed between two jaws of untarnishable metal. The edges are bevelled on the side which faces the lens and are ground perfectly sharp and straight. The method of mounting the jaws varies in different designs; but there is invariably provision for opening or closing the slit by means of a screw of fine pitch, which is usually provided with a divided drum so that the slit width can be set at definite values. In the majority of cases one of the jaws is fixed, and the slit opens and shuts unsymmetrically. In some of the more expensive slits both jaws move simultaneously, so that the centre of the slit remains in the same position whatever the slit width. Some workers have a preference for symmetrical slits; but except when the instrument is used as a monochromatic illuminator, it is doubtful if there is appreciable advantage in their use. On the other hand, with an observer who is apt to take an instrument on trust, they may be a source of danger. When precise measurements are in question it must never be assumed that the centre of the slit has exactly the same position for different widths. It is not therefore satisfactory to vary the slit width when dealing with spectrum lines of different brightnesses as advocated by some writers. If some lines are too bright with the slit width necessitated by the faintness of others, their brightness should be diminished by placing absorbing glasses between the slit and the source, removing these when working on the faint lines. There is no slit mechanism sufficiently perfect to enable one to assume an absolutely symmetrical opening and closing of the slit to the very high degree of accuracy of which a well-designed spectrometer is capable.

Whether symmetrical or unsymmetrical, a good slit will also have provision whereby the exact parallelism of the jaws may be adjusted. In certain classes of work extremely fine slit widths, as little as 0·01 mm. for instance, are used. It is only with the most accurate adjustment for parallelism that slits of such fineness are possible.

When a good slit has been obtained it should be treated with the utmost care. To close the jaws completely is almost invariably fatal, and it is rare to find a slit which has been in general laboratory use for any length of time which will close down to a very narrow width without showing streaks across the spectrum due to some parts of the jaws being in contact.

of the instrument and the errors to which it is liable, and in devising means for their elimination or correction. The proper attitude to adopt to any instrument is one of distrust and suspicion; and it should not be regarded as accurate in any particular until it has been carefully verified by tests as many and varied as possible.

For working to anything less than ten seconds verniers are unsatisfactory, and the scale has to be read by two or four micrometer microscopes. The need for an even number of microscopes arises from the very probable existence of "centering error." This error is due to the scale not being exactly concentric with the axis of rotation.

Suppose C_1, *Fig.* 10 (*a*), is the centre of the scale, and C_2 is the axis about which the moving portion of the instrument rotates. Let C_2V be the arm carrying the vernier or micrometer. Then the reading will be at R, where C_2V intersects the scale. Let the line of excentricity EE_1 make an angle θ_0 from the zero point Z of the scale and let $C_1C_2 = e$. It is clear that if the scale reading is θ_0, that is, when the radius bar lies along EE_1, the reading is independent of the displacement e, and is therefore the same as if e were zero. When the bar is in any other position the angle through which it has been rotated from EE_1 is $R\hat{C}_2E_1$. But the angle recorded on the scale is $R\hat{C}_1E_1$. The recorded angle is therefore too small in the case shown by $C_1\hat{R}C_2 = e \sin \theta/r$, r being the radius. For any scale reading ϕ, therefore, there is an error of $-(e/r)\sin(\phi - \theta_0)$. This is of maximum value when the vernier arm is perpendicular to EE_1, and is zero when it is parallel to EE_1. The curve of error will be a sine curve of the form shown in *Fig.* 10 (*b*). The actual error between two readings ϕ_1 and ϕ_2 is the algebraic difference between the ordinates of the curve at these readings. It may therefore vary from zero to $2(e/r)$ depending on the values of ϕ_1 and ϕ_2.

We notice, however, that the ordinate at any point of the curve is equal and opposite to that at a point 180° from it. Thus if there are two verniers 180° apart the errors of one will always be equal and opposite to those of the other, and the mean of their readings will give the true angle.

Thus centering error can be eliminated if the mean reading of two verniers (or micrometers) 180° apart are taken. Spectrometers, except of the most elementary type, are therefore fitted with two or four verniers (or micrometers) equally spaced round the circle.

This, however, while it eliminates centering error where the direction and magnitude of the eccentricity is constant, does not eliminate every possibility of error from defective centres. If in the course of rotation the relative position of C_1 and C_2 alters, which is certain to happen to a greater or less extent unless the bearings are perfect in the literal sense of the term, the errors in each vernier reading need not follow the simple sine law

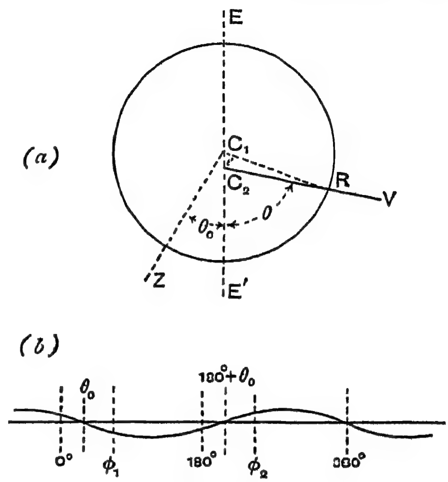


FIG. 10.

and will not necessarily be eliminated by taking the mean of readings 180° apart. An instrument, even if perfect when new, is certain to develop erratic centering errors of this type in the course of time, unless the rotating portion is very carefully balanced so as to press uniformly at all points of the bearing. If the contact is heavier on one side than another, the axis will be worn slightly elliptical in use, work being more concentrated in some ranges of angle than in others, and therefore e will change both in magnitude and direction for different readings.

Apart from centering errors, the scale divisions may themselves be non-uniform. There may be gradual variations of the size of a division due to unsuspected centering errors in the dividing engine with which the graduations were cut, or there may be irregular displacements of the division lines owing to imperfect working of the engine.

To calibrate the circle for every division is a very tedious business, and is not at all easy to perform with great accuracy except where special apparatus exists for such tests. On this account it is usual in the best instruments to provide means for altering the orientation of the scale with respect to the verniers or microscopes, in order that a particular measurement may be repeated, using different graduations distributed round the scale. In this way, if sufficient measurements are made, the graduation errors tend to average out.

The foregoing considerations apply both to

the measurements of the rotation of the telescope and rotation of the table. A serious defect in the design of most spectrometers is that the axis of rotation for the table is much less satisfactory than for the telescope; consequently measurements of table rotations, even when made on the same circle as those of the telescope rotations, are liable to much greater errors due to defective action of the centre. This is a serious matter in the general use of such instruments where, for certain purposes, such as the measurement of prism angles, for instance, it is necessary that the table rotation should be measurable with the very greatest accuracy.

There is one more important point about the centering arrangements of the spectrometer. The two moving parts should be mounted on a single axis, the table being, of course, uppermost; and the clamping devices and fine motion fittings should be so arranged that there is no frictional connection whatever between the two rotatable parts. If this is not secured the usual result is that the part, which in a given experiment is supposed to remain fixed, suffers a slight drag by the moving part in the direction in which the latter is moved.

(iv.) *A Standard Spectrometer.*—In *Fig.* 11 is shown a spectrometer designed in its essentials by the writer and in detail by Messrs. E. R. Watts, Ltd., of London, who made the instrument for the Optics Department of the National Physical Laboratory. The diameter of the base is 30 inches, which gives the scale of the figure.

The aim in the design was to avoid all sources of error which could be foreseen and to render easy the detection and elimination of such unforeseen and unavoidable errors as might still be present in the completed instrument.

The base is a circular iron plate, strengthened with radial and concentric webs on the lower side, and provided with two concentric slots of ⊥ section in the upper face. These are shown at 1 in the figure. The collimator, 2, is of 3½-inch aperture, and can be clamped to the base in any azimuth by four bolts, the heads of which are in the broad portion of the ⊥ slots. The telescope, 3, of 2½-inch aperture is mounted on an accurately balanced arm. In order to reduce the weight on the centres to a minimum, light alloys were employed wherever mechanical considerations permitted. A second telescope, 4, the uses of which are indicated later, is provided. This telescope is identical with 3, except that it stands on a pillar which can be clamped in any position round the base in the same way as the collimator. When not required it can be removed. Both telescopes and collimator are provided with adjustments for horizontality of axis, which are carefully designed to give great rigidity without the introduction of strain in the tubes.

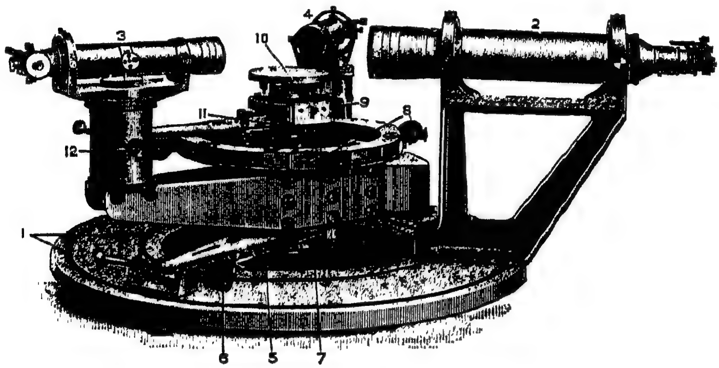


FIG. 11.

The centre consists of a strong tapering pivot. About the lower portion of this the telescope rotates, fine adjustment being obtained by means of the clamping collar, 5, and tangent screw, 6. A ring, 7, is screwed on a thread at the lower end of the pivot. This ring supports a ball-bearing, which in turn supports the weight of the moving telescope system. By adjusting 7, the moving system can be slightly raised or lowered to vary the tightness of the centre and secure proper freedom of rotation.

On the upper half of the central pivot, a second moving system, consisting of the table and graduated circle, rotates. The circle, which is of 15 inches diameter, is completely enclosed by a dust-proof cover. The scale is read by four micrometer microscopes, 8, reading directly to half-seconds and by estimation to less. These are attached

to a ring associated with the lower moving system and can be fixed in any position relative to it. The table itself can rotate for adjustment purposes independently of the circle. For this motion it is actuated by an endless tangent screw, 9, which can be thrown in or out of gear by a paul, 10.

When using the instrument for purposes in which rotation of the telescope is to be measured, the upper moving system has to be clamped to the pivot, thus fixing the position of the circle. The clamping device employed is such that no radial force whatever is exerted on the system. There is therefore no possibility of centering error being introduced by the process of clamping. When measuring rotations of the table, the telescope being fixed, the upper moving part is unclamped from the centre. Fine motion of the circle relative to the telescope is obtained by means of the clamping collar, 11, and tangent screw, 12. This collar can be removed entirely when not required, thereby removing any danger of a slight frictional drag on the circle and table when rotating the telescope. If desired, the tangent screw, 12, can be transferred to the support of the additional telescope, 4.

This instrument, it will be seen, can be used with equal accuracy for all purposes for which a spectrometer can be employed.

§ (7) The Prism.—In the last section the principal features and requirements of a spectrometer designed for the most accurate refractometry (and goniometry generally) were indicated. We shall now look at the requirements of the prism on which such measurements are to be made. In the first place, we must decide to what angle it should be cut. In discussing the sensitivity of the two principal methods available for determining the index of the prism, viz. the minimum deviation and grazing incidence methods, we saw that the sensitivity considerations alone would lead us to employ the largest angle for which light would still be transmitted by the prism. Further, from *Fig.* 4, the extent to which an error in the angle of the prism affects the result is seen to be nearly the same for all prism angles for the minimum deviation method, while it diminishes appreciably as the angle increases in the case of grazing incidence. This would also lead us to use as large an angle as possible.

There are other considerations, however. It is clear that the more nearly grazing the emergent ray is, the narrower is the transverse aperture of the beam which enters the telescope. This results in loss of definition. Up to a certain extent this is found to be without effect on the precision of the settings; but when carried too far some accuracy is lost, particularly in the method of grazing incidence, in which, as we shall see later, the setting is unsymmetrical. This consideration would lead us to use small angles. The final consideration, and the decisive one in the present stage of angular metrology, is that angles which are submultiples of 360°, such as 30°, 45°, 60°, and 90°, can easily and conveniently be measured with greater accuracy than those of intermediate value.[1] This fact leaves the balance of advantage with the 60° prism. Not only so, but if the substance is isotropic, or if it is a uniaxial crystal with its axis perpendicular to the principal plane of the prism, so that it has the same refractive index in any direction in the principal plane, it is possible, if all three faces are polished, to make measurements with each angle in turn. Glazebrook has shown[2] that if the individual angles of an approximately equilateral prism do not differ too much from 60°, the mean of the deviations when each angle is used in turn is equal to the deviation corresponding to a prism angle of exactly 60°. It must be borne in mind, however, that this result assumes that the prism is a true prism, *i.e* the three edges are supposed to be accurately parallel. By the assumptions on which the formulae of prism refraction are based, the angle of the prism is the angle between the two plane faces at which refraction occurs. If the prism is pyramidal, the sum of the angles between the faces taken in pairs exceeds 180°, and the mean deviation corresponds not to 60° but to $60°+\xi/3$, where the sum of the three angles is $180°+\xi$. It is shown in the article on "Goniometry," where the methods of measuring prism angles are dealt with, that for a 60° prism, $\xi=0{\cdot}279P^2\times10^{-5}$ seconds, where P seconds is the inclination of an edge of the prism to the opposite face. Thus if the pyramidal error of the prism is 10 minutes, the sum of the angles is 180° + 1 second.

While it is necessary to bear this in mind, and to test for pyramidal error, no prism which is otherwise good enough for precision determinations is likely to have more than a few minutes' pyramidal error at the outside, in which case the excess of the angles over 180° can be neglected.

Thus with a 60° prism it is possible to dispense altogether with a precise knowledge of the angles, thereby eliminating about half of the possible experimental error of the index determinations.

A prism which is to be used for determinations of great accuracy must be in the highest degree homogeneous and have its surfaces very accurately plane.

The total effect of defects of surface and

[1] See article on "Goniometry," § (3).
[2] See footnote to paper by Gifford, *Roy. Soc. Proc.*, 1902, xx. 230.

non-homogeneity on the transmitted wave-front, and also the planeness of the surfaces themselves, can be tested by interference methods (*vide* "Interferometers: Technical Applications," § (4)), and from these results the departures from homogeneity of the material can be deduced. The great skill with which surfaces can now be polished renders it easy to get prisms with surfaces of almost perfect flatness; but it is by no means easy to obtain a prism of satisfactory homogeneity. In fact the homogeneity of the material is what sets the limit at present to the accuracy to which it is worth while attempting to push refractometric measurements.

Theoretically, by increasing the mechanical perfection of the spectrometer, and by using telescopes of greater aperture and power, the precision of the angle measurements could be increased very considerably beyond what is necessary for fifth place accuracy in the index measurements; but the useful size of the telescopes is limited by the size of prism which can be obtained of satisfactory homogeneity. It is of little use at present to attempt to use prisms of more than 2-inch side.

§ (8) Adjustment of Spectrometer.—There are two principal adjustments which have to be made before a spectrometer is ready for use. The collimator and telescope have to be focussed for parallel light, and they have also to be adjusted so that their optic axes are perpendicular to the axis of rotation of the instrument. There are various ways of performing these adjustments; but by far the most convenient makes use, in each case, of an auxiliary telescope. Since such an auxiliary telescope is of the greatest convenience in many methods of using the spectroscope, particularly for goniometry (*vide* "Goniometry," § (2)), the experimenter should always provide one for his instrument. The telescope should be similar to that of the spectrometer, though it need not have a micrometer eyepiece, and should be mounted on a rigid stand with adjustable feet as described in the article just quoted.

The adjustment for collimation is very simple. First one telescope and then the other is focussed on the slit of the collimator. The two telescopes are then placed in line with each other, and the cross-lines of one of them illuminated by means of a lamp placed behind the eyepiece. If on looking through the other the cross-lines of the first appear in sharp focus, the collimation is correct. If, however, the collimator was not in correct adjustment to start with, the telescopes would each be focussed for an object either within or beyond infinity. Suppose the latter, *i.e.* the telescope cross-lines, are nearer the objective than the true focus, then if the light passes in the reverse direction, rays from a point in the plane of the cross-lines will leave the objective as a divergent beam. Thus when the first telescope is employed as a collimator with respect to the second, the rays from it *diverge*, whereas the second telescope is focussed for *converging* rays, and so the cross-lines of the first are not seen in sharp focus. Half the adjustment necessary to bring them into focus should be made with *each* telescope, after which both will be correctly collimated. Either can then be directed towards the collimator and the latter adjusted to bring the slit into focus.

If the initial error of collimation is great, the process should be repeated; a very few repetitions will secure perfect adjustment.

Apart from spectroscopic work, any three telescopes or collimators can be focussed for infinity in a very short space of time by adjusting them in this way until they will focus on each other in pairs.

In the case of the spectroscope the collimation must be adjusted for the wave-length of the light with which the instrument is about to be used, since the chromatic aberration of the lenses is always appreciable near the extremities of the spectrum.

The adjustment of the optic axes of the telescope and collimator to be perpendicular to the axis of rotation of the former is also most conveniently performed with the aid of the auxiliary telescope. The method is as follows:

The rotating telescope is turned to an oblique position, such as T_1, *Fig.* 12, and a prism is set on the table and adjusted so that one of its faces reflects the light from the collimator into the telescope. It is necessary to mark the approximate centre of the slit. This is usually done by stretching a fine wire or fibre across it; but it is much better to employ two such fibres, separated by about 0·5 to 1 mm. The point half-way between these, which can be judged with ample accuracy, is taken as the centre of the slit. This obviates making settings where the slit is crossed by the fibre. The prism is levelled by means of one of the adjustment screws of the table so that the centre of the slit comes on the cross-lines of the telescope. The latter is now rotated out of the way and the auxiliary telescope, T_2, set up in its place. The latter is adjusted by means of its own levelling screws until the centre of the slit is exactly on the cross-lines. It will be clear that by this process we have placed the auxiliary telescope with its axis in exactly the same direction as that of the telescope T_1 in its initial position. The prism is now removed, and T_1 is swung into a position in line with T_2. A lamp is placed behind the eyepiece

of the latter to illuminate the cross-lines, which are then visible in the field of T_1. If the adjustment is correct, so that the axis of T_1 generates

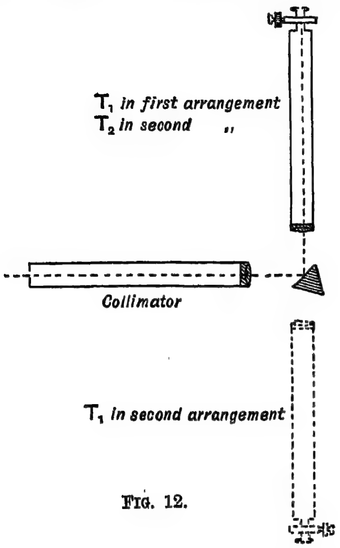


FIG. 12.

a plane on rotation, it is evident that the cross-lines of T_2 and those of T_1 itself will be at the same height in the field. But if the optic axis of T_1 is not perpendicular to the axis of rotation, it generates a cone, and in this case there is a vertical displacement between the cross-lines of T_2 and those of T_1. Obviously the angular extent of this displacement is twice the amount by which the axis of T_1 departs from perpendicularity to the axis of rotation. T_1 is therefore adjusted to reduce the vertical displacement by one-half, when its axis should be in correct adjustment. It is then brought into line with the collimator and the latter adjusted with respect to it. It is advisable to repeat the process if the initial error is large.

This method of axis adjustment can be performed quite quickly in practice, and is much less tedious and also more accurate than the customary spirit-level methods.

§ (9) THE USE OF THE SPECTROMETER FOR REFRACTOMETRY. MINIMUM DEVIATION METHOD.—The prism of which it is intended to measure the refractive index is placed on the table. Let F_1 and F_2, *Fig.* 13 (*a*), be the faces which are to be employed in the measurement; one of them, F_1, should be placed approximately perpendicular to the line *ab* joining two of the adjustment screws of the table. The telescope is put in any convenient position, say T, *Fig.* 13 (*b*), and the table rotated so that an image of the slit is seen by reflection in F_1. The centre is brought to the level of the cross-lines by the screw *a*, *Fig.* 13 (*a*). The table is then rotated until F_2 is in the position previously occupied by F_1, and the height of the image adjusted by the adjustment screw *c*. Since this does not affect the previous adjustment of F_1, both F_1 and F_2, and, therefore, also the edge in which they intersect, are now parallel to the axis of rotation. The prism is then in adjustment for making the index measurements. The minimum deviation method is carried out as follows:

The prism table is rotated so that the light transmitted by the prism is deviated to one side as in *Fig.* 13 (*c*). The table is rotated, and the image followed with the telescope until the deviation is a minimum. Many observers use special methods for determining when the deviation is a minimum; but it is quite easy to obtain the correct position without them. In case of doubt, however, the best method is to note the two positions of the table for which the deviation has a particular value just a little in excess of the minimum, and then set it half-way between these positions. If the telescope is provided with a micrometer eyepiece the best method of determining the position of the telescope is to clamp it in the approximately correct

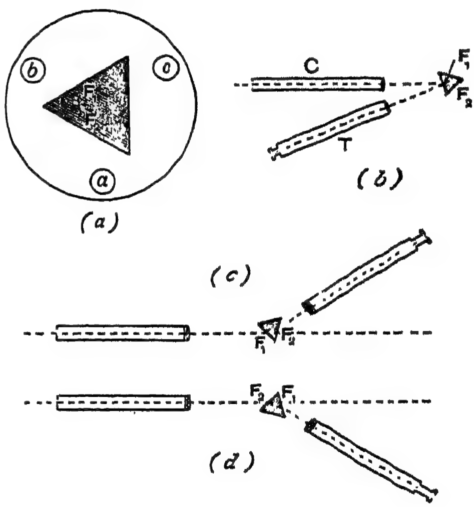


FIG. 13.

position and note the scale reading. This should be left fixed and the series of actual settings on the spectrum line made by means of the micrometer eyepiece. The individual settings will usually be of greater precision, and consequently fewer will be required to attain a given accuracy of the mean than if the telescope is moved bodily at each setting.

Having determined the direction of the refracted ray in this position, the prism table

is rotated so that the deviation is to the other side, *Fig.* 13 (*d*), and the same process of measurement is performed. The difference in the scale readings for the two cases, corrected for the difference in the micrometer means, is twice the angle of minimum deviation.

The necessity of doubling the angle by employing the two positions of *Fig.* 13 (*c*) and (*d*) arises from the difficulty of taking the zero reading with the telescope and collimator in line. These are focussed up for the wave-length of the monochromatic radiation for which the determination is being made. If a direct zero reading were attempted the light in the image of the slit would comprise all the components emitted by the source, and in general the settings could not be made without readjustment of focus.

Such, in outline, is the minimum deviation method. There are various matters of technique which must receive attention, however, if the highest accuracy is aimed at in the results.

(i.) *Position of Prism on Table.*—This must be so adjusted that in both positions, *Fig.* 13 (*c*) and (*d*), the light comes from the same area of the collimator lens and enters the same area of the telescope lens. If, for instance, the prism is arranged with its centre coincident with the axis of rotation, this condition will not be fulfilled. Thus in *Fig.* 14 the portion of the apertures used is mostly to one side of the axis in (*a*) and to the other side in (*b*). The errors which may arise from such an arrangement, on account of imperfect focussing or spherical aberration, are discussed in another article.[1] The prism must be moved towards its base so that it employs the apertures centrally, as in *Fig.* 14 (*c*). To make this adjustment, examine the exit pupil of the telescope with a magnifier. The illuminated aperture of the prism will be seen within the circular aperture of the telescope. Move the prism until its illuminated face appears symmetrical with respect to the telescope lens, taking care to keep the bisector of the refracting angle above the axis of rotation. When this is done the prism aperture will be symmetrical with respect both to the telescope and collimator axes in both positions, *Fig.* 13 (*c*) and (*d*). Needless to say, the parallelism of the refracting edge to the axis of rotation should be checked after the prism is in its final position.

(ii.) *Adjustment of the Light.*—The source of light has to be adjusted so that the aperture of the collimator is evenly filled with light. With broad sources such as flames it is sometimes good enough simply to place them at a little distance from the slit. With vacuum tubes or other small sources it is essential to form an image of the source on the slit. The source should be placed at some distance from the slit and adjusted until the rectangular

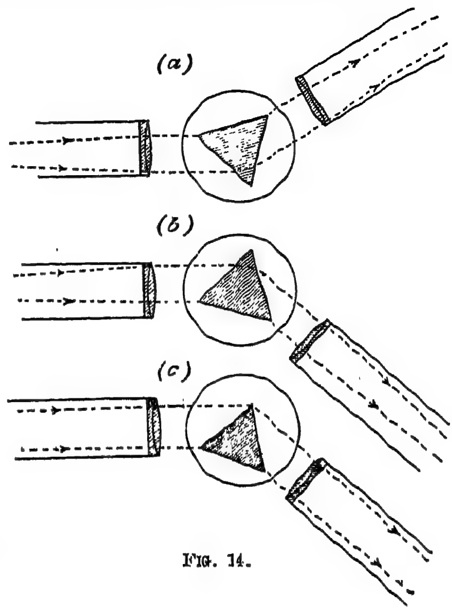


FIG. 14.

patch of light, which will be seen falling on the collimator lens, is at the centre of the latter. A lens is then interposed and adjusted until an image of the source is formed on the slit. The aperture and focal length of the lens should be such that the cone of rays which it concentrates on the slit just fills the lens of the collimator. If a wider cone is supplied some of the light is scattered from the sides of the collimator tube and may ultimately reach the field of view, adding to the stray light which is always to be found there.

(iii.) *Illumination of the Cross-lines.*—It is necessary in order to make comfortable and accurate settings that the cross-lines should be clearly seen. Further, it is necessary, particularly towards the blue end of the spectrum, to illuminate them with light of the same wave-length as the spectrum line on which measurements are being made. If this is not done very serious troubles arise, due to chromatic parallax, which make accurate settings an impossibility. The peculiarities of chromatic parallax, which arises from the chromatic aberration of the eye and the restricted size of the exit pupil of the telescope, are explained in another article.[2]

There are several methods of eliminating chromatic parallax,[3] but the most convenient for refractometric work, and in fact for any

[1] "Goniometry," § (1).

[2] "Eye, The" § (27).

[3] Guild, *Proc. Phys. Soc.*, 1917, xxix. 311.

spectroscopic work except when there are numbers of lines closely packed together, is to use a wide slit, say 1 to 2 mm., with a fibre, similar to those used for the cross-lines, stretched along its centre. Each spectrum line then gives rise to a band of light with a black line down the middle. The band serves as a background for the cross-lines, and settings are made on the central line. Suitable fibres of quartz or glass can be mounted on large washers, together with a horizontal pair to mark the centre of the slit, and protected with microscope cover glasses. These can then be affixed in front of any spectroscope slit with wax. It will be found useful to prepare a number of such fibres of varying thickness, say from 5 to 10 thousandths of a millimetre, and employ that which gives the greatest comfort under the circumstances of the observation. The collimation has of course to be adjusted with the fibre in position.

This arrangement has several advantages in addition to the elimination of chromatic parallax. Every worker with a spectroscope has realised the difficulties of setting cross-wires accurately on a fine spectrum line. Even when the general field illumination is satisfactory and the cross-lines are distinctly visible there is an annoying uncertainty as to the exact point of superposition. Diffraction effects make themselves evident in a wavering and loss of definition at the point of intersection, which is very trying. With the arrangement described there are no such effects at the moment of contact. Everything appears as distinct and clear as if it were a diagram drawn on paper. Secondly, the tiresome fluctuations of accommodation, to which the eye is subject when using faintly illuminated cross-lines, are practically absent. Thirdly, the full brightness of the line is utilised. With a fine spectrum line the field illumination has to be low enough to give a satisfactory contrast. With faint lines this involves working at illuminations for which the acuity of vision is not great. With the wide slit method the full brightness of the line is utilised as background for the cross-lines, so that the maximum use is made of the available light.

The method can be used with advantage for all cases in which there are not several lines packed close together. With sodium light, for instance, the slit images for D_1 and D_2 are superposed, but the central line of each is quite clearly defined. It is only when the bands of *several* close lines overlap that the individual images of the fibre become difficult to see.

(iv.) *Elimination of Scale Errors.*—When determining the scale reading by means of the micrometer microscopes, settings should be made on several consecutive rulings. This tends to reduce errors due to local irregularities. To eliminate progressive errors of ruling it is necessary to repeat the whole determination, say half-a-dozen times, with the scale shifted through about 60° in relation to the microscopes each time.

(v.) *Careful Manipulation.*—In general the greatest care in handling the instrument in the course of a series of observations is essential. No spectrometer will give consistent readings if subjected to strains and flexures. In particular the rotation of the telescope from one position to another must be made slowly and with the careful avoidance of sudden accelerations; and the manipulation of the micrometer eyepiece must be performed without any tendency to push or pull the eyepiece in the direction of measurement.

§ (10) CRITICAL ANGLE METHOD.—The essential condition for this method is that a converging beam of light should fall tangentially on one face of the prism, *Fig.* 15. The

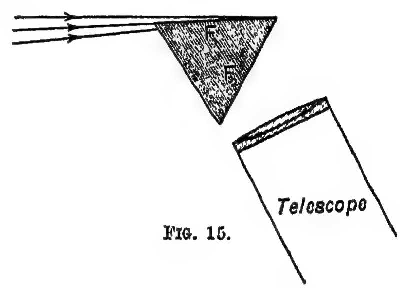


FIG. 15.

rays are refracted into the prism and fall obliquely on the second face, at which they are refracted into the air and may be received by a telescope suitably placed. It is clear that all rays which have the same angle of incidence will be deviated equally by the prism and will combine in the focal plane of the telescope to form a fine line parallel to the refracting edge of the prism. Actually the line will be slightly curved (as in the case of the image of a slit) on account of the fact that the deviation for rays which are inclined to the principal plane of the prism is greater than for rays in the principal plane. In practice, if the adjustments of the prism have been properly carried out, we are only concerned with rays in the principal plane.

We have, then, a fine line in the field of the telescope corresponding to each angle of incidence present in the converging beam. The resulting effect will be a band of light. It is evident that the incidence angle is sharply delimited at 90°, when the rays are just grazing. There are therefore no rays for which the angle of refraction is greater than for those at grazing incidence, and the band of

light in the telescope must terminate abruptly at the line corresponding to these rays. Thus the band will have a sharp edge; and if the cross-lines are set on this edge the inclination of the telescope axis to the normal of the face F_2 will be the angle of emergence, e, in the formulae which were developed in § (2) for the case of grazing incidence.

To make a determination by this method, the prism is mounted on the spectrometer table and levelled as described in the preceding paragraph. The source S is placed in line with one of the faces, *Fig.* 16 (*a*), and a lens L inserted to produce a converging beam. The distances of S and L should be such that an image of S is formed about an inch or so beyond the prism.

The telescope is rotated until the critical edge is on the cross-lines and a series of settings made. As in the case of minimum deviation, it is preferable to make these settings with the micrometer eyepiece, leaving the telescope clamped in one position throughout.

Having determined the reading for the critical edge in the position shown, the source and lens are shifted to S′ and L′ so as to bring the illumination from the other side of the prism. The telescope has then to be moved to the dotted position and the settings repeated. If ϕ is the angle between the two positions of the telescope, and e is the angle of emergence (the sign of e being given in accordance with the convention adopted in § (2)),

$$2e = \phi + \hat{C} - 180^\circ.$$

It has to be noted that the angle of emergence in the first position corresponds to a refracting angle $\hat{A}$, while in the second position the refracting angle is $\hat{B}$. If these differ the angles of emergence are not equal in the two cases. However, if they do not differ greatly, the value of e given by $\frac{1}{2}(\phi + \hat{C} - 180^\circ)$ will correspond to the mean of the angles $\hat{A}$ and $\hat{B}$. Clearly, if the measurements are made with the light incident on each of the faces in turn, giving three nearly equal values, ϕ', ϕ'', ϕ''', for the rotation of the telescope, we can put $e = \frac{1}{2}\{(\phi' + \phi'' + \phi''')/3 + 60^\circ - 180^\circ\}$; and the value of e so obtained will correspond to a prism angle of exactly 60° (assuming no pyramidal error).

The order of magnitude of ϕ is about 60°, more or less. This method, therefore, involves the measurement of a fairly large angle. An alternative arrangement, which only involves measuring a small angle, is shown in *Fig* 16 (*b*). It requires the use of the additional telescope on a separate stand, to which reference has already been made (§ (8), p. 765). The light is incident along the two faces which contain the refracting angle. The telescope T_1 is rotated so as to receive one of the emergent beams. The auxiliary telescope T_2 is placed so as to receive the other. The method of making the measurement is as

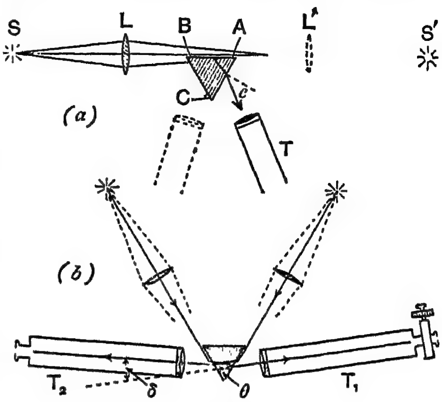


FIG. 16.

follows: The critical edge in the field of the fixed telescope T_2 is brought on the cross-lines by means of the tangent screw which rotates the table; then the telescope T_1 (or its micrometer eyepiece) is adjusted so that its cross-lines are set on the other critical edge. *Both* settings should be repeated several times, the telescope reading being noted each time. If δ is the angle between the axes of the two telescopes when they are both correctly set on the respective critical edges, $2e = \delta - \theta$. If the process is repeated with each angle of the prism in turn, giving three nearly equal values, δ', δ'', δ''', then $2e = (\delta' + \delta'' + \delta''')/3 - 60^\circ$, where e is the angle of emergence corresponding to a prism angle of 60°.

To measure δ we illuminate the cross-lines of T_2 and treat it as a collimator with respect to T_1, which should be rotated until the two sets of cross-lines coincide. The angle through which it has to be moved is the angle δ.

In order that all the observations may be made without removing the prism from the table, it is convenient to arrange the height of the latter (or of the telescopes, if this is more convenient) so that the lower half of the telescope objectives are utilised for the critical angle measurements, while the upper half are above the level of the prism top and are available for the collinear measurements.

As regards the latter there are one or two precautions to be observed. In the first place, it is not satisfactory to attempt to set the two cross-lines in exact coincidence. Such a setting is very insensitive. The telescope axes should be very slightly out of line in the *vertical* direction, so that one cross is a little above or below the other. Then when the axes are in line horizontally the two sets

of wires form a small symmetrical rhomboid. If a suitable vertical displacement is chosen, this rhomboidal setting can be made with great precision. In repeating the determinations *ab initio*, the cross-lines, which in the first case were lower, should be the higher next time, and so on. In this way any error due to want of exact parallelism of the two sets of wires is eliminated, and a true setting corresponding to exact coincidence of the axes is secured.

As in the case of the minimum deviation method, the final accuracy of the result depends on the care with which the manipulation is performed, and on the arrangement of the apparatus so that the processes involved in the actual measurements are rendered as few and as simple as possible. In the critical angle method the success or failure largely depends on the illumination arrangements. If these are satisfactory the method is easy and accurate; but if they are otherwise, as may very easily be the case, or if they require readjustment in the course of a determination, the greatest trouble may be encountered. It is undesirable, for instance, to have to shift the source so as to illuminate first one face of the prism and then the other. Two sources must be set up to begin with; or, what is better still, an arrangement should be adopted by which one source may be employed to illuminate both faces simultaneously.

A convenient and easily constructed illuminator for this purpose is shown in plan in *Fig.* 17. D is a hole of about 3 mm. diameter a little outside the focus of a condensing lens L. The source, if it has appreciable width, is placed behind D, or an image of it is formed at D by another lens, as if D were the slit of a spectroscope. The light leaves the lens L in a slightly converging beam. It then meets a 60° prism, which should be of clear colourless glass polished on all three sides, but which need not otherwise be of good quality. This splits the beam, by internal reflection, into two beams which emerge in the directions shown. These beams, when they have separated to an extent of 15 cm. or so, meet a pair of right-angled prisms so placed as to make them reconverge at an angle of 60°. The various parts can be mounted permanently on a single stand. The supports of the right-angled prisms should be capable of rotation, as this facilitates adjustment.

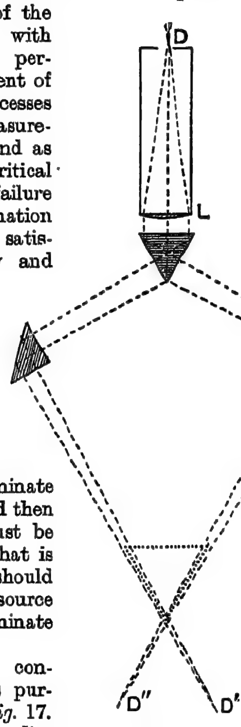


FIG. 17.

The illuminator is so placed in relation to the spectrometer that the prism under test occupies the position which is shown dotted in the figure.[1] The lens L should have a power of about 10 diopters, and the distance of D from it should be such that images of D are formed at D′ and D″, about two inches beyond the apex of the prism. When the illuminator has been placed approximately in the correct position, a slight rotation of the right-angled prisms will usually be sufficient to give satisfactory bands with sharply defined edges.

If the source is rich in light of other wave-lengths than that for which the measurement is being made, stray light of these wave-lengths almost always appears in the field, and may give rise to chromatic parallax troubles. To overcome this difficulty filters must be used in front of D to cut out or reduce the objectionable rays.

In determining δ, the suitable illumination of the cross-lines of T_2 may present some difficulty. It is evident that when the telescopes are sharply focussed on the critical edge of the refracted beams they are collimated for the wave-length of the light employed. In the presence of the usual amounts of chromatic aberration, they will not be in correct focus for other wave-lengths. If the refracted light is in the yellow or green part of the spectrum it will generally be quite satisfactory to put a table lamp behind the eyepiece of T_2 with, possibly, the interposition of a colour filter to approximate the colour to that for which the telescopes are focussed. If, however, the refracted light is not near the middle of the spectrum it is not usually possible to get sharp enough focus with colour filters, and it is necessary to employ light of the wave-length for which the index determination is being made. An arrangement which proves convenient, except with faint lines, is to mount a small mirror close up to the eyepiece of T_2 and adjust it so that it sends the refracted light which emerges from the eyepiece back into it to illuminate the cross-lines. This adjustment has to be made with some nicety,

[1] Particulars of this method are given in a paper by Guild and Dale, *Trans. Optical Soc.*, 1920, xxii. pt. 3.

so it is expedient to mount the mirror on the end of a lever, pivoted on a suitable stand. It can then be swung out of position when it is desired to look into the eyepiece of T_2, and swung back to position for the collinear measurement, without upsetting its adjustment. In this way the difficulty of securing light of the same wave-length for all the measurements involved can be overcome.

Special Case of $\mu = 1{\cdot}53$.—From equation (4), § (2), if $\mu_2 = 1{\cdot}528$ the emergent angle is $-30°$, and the two critical rays are collinear, provided $\theta = 60°$ exactly. Thus for indices near this value δ is very small, and can be measured on the micrometer eyepiece without rotation of the telescope at all. For such indices we thus have a useful check on the accuracy of more general methods; since we eliminate completely the axis of the instrument and any undetected peculiarities it may possess.

It is this feature of the method which invests it with importance in precise refractometry. It is not in itself quite so easy to perform as the minimum deviation method, and is not suitable at all except for bright spectrum lines; but the fact that it can be made to involve only very small rotations of the telescope or, in a particular case, no rotation at all, makes it highly useful as a check to be used in conjunction with other methods which depend for their accuracy on the accurate measurement of comparatively large rotations.

It is desirable therefore to investigate the errors to which the results may be liable on account of the unsymmetrical nature of the setting. There are two points which require attention. In the first place, on account of the phenomenon of irradiation, the bright area will encroach on the dark area and the edge will be displaced on this account. The magnitude of this error can easily be determined experimentally. As the observer should always determine it for himself with the telescope to be used for the measurement, a method of doing so may be described.

A prism P, *Fig.* 18, is placed at the focus of a long focus telescope objective L with one of its faces approximately at 45° to the axis. A glass plate A is placed a little behind the prism, also inclined at 45° to the axis. A source of light S_1 and a sheet of ground glass G_1 are placed so as to illuminate the face of the prism, and a second source S_2 and ground glass G_2 are placed where they will illuminate the plate A. The observing telescope is placed at T and focussed sharply on the edge E of the prism. If S_1 is turned on and S_2 is off, the field will be bright on one side of E and dark on the other; while if S_2 is on and S_1 is off, the bright and dark sides are reversed. If, then, the cross-lines are set on the edge when first one side and then the other is bright, the difference in the micrometer readings will be twice the error due to irradiation. When such measurements are made it is found that the error amounts to about 40 seconds divided by the magnification of the telescope. Thus for a telescope of power 20 the error is only about 2″. Further, this is found to be independent of the brightness over a very wide range, and also of the colour. Thus one may find very precisely the correction to be applied to critical angle readings on account of irradiation. It will be clear, of course, that this correction includes any other "personal equation" error which may influence the setting in addition to irradiation.

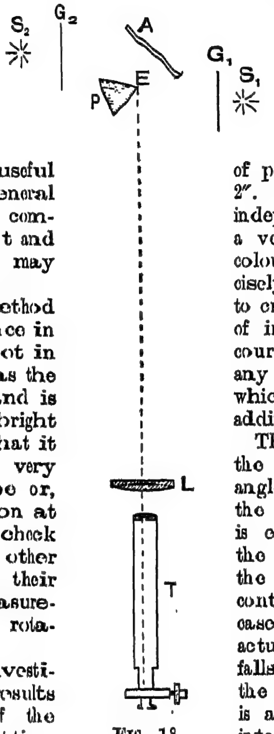


FIG. 18.

There is one particular in which the conditions in the actual critical angle setting differ from those in the arrangement just described. It is evident that in the latter case the intensity is uniform right up to the edge, and that a sudden discontinuity occurs there. In the case of the critical edge there is no actual discontinuity. The intensity falls off more and more rapidly as the critical angle is approached, and is actually zero at that point. The intensity curve can be calculated from Fresnel's equations for refraction and reflection, correction being made for the fact that the rays in the refracted beam are compressed more and more closely as the critical angle is approached. These curves have been calculated by Kruess,[1] and in *Fig.* 19, the intensity curve for a band 1° wide is given for $\mu = 1{\cdot}6$, and $\mu = 1{\cdot}05$. What will the eye adopt as the "edge" of such a band? There is reason for believing that the "edge" will be that part at which the intensity is diminishing most rapidly. While it is not evident from *Fig.* 19, on account of the scale on which it is drawn, the slope of the intensity curve increases continuously

[1] Hugo Kruess, *Zeit. f. Instk.* xxxix., March 1919, p. 73.

up to the critical angle, and is therefore greatest at this point. Therefore, if the above supposition is correct, the same setting should be made as if the brightness were uniform up to the edge and suddenly dropped to zero.

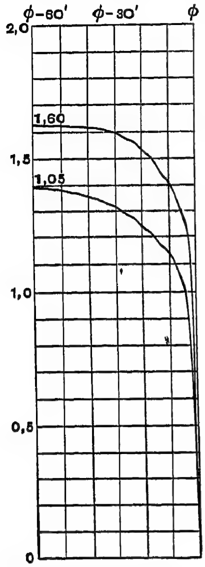


FIG. 19.—Intensity in neighbourhood of Critical Angle. ϕ=Critical Angle.

That this somewhat unexpected result is correct to a high order of accuracy has been demonstrated experimentally.[1] Consequently the critical angle setting only requires the easily determined correction, mentioned above, for irradiation and personal equation. The results of the method may therefore be employed with confidence for comparison with those obtained by the minimum deviation method.

III. REFRACTOMETERS

§ (11).—The methods which have been described in some detail in the preceding paragraphs are those suitable for the most precise absolute determinations. They involve a considerable amount of time to carry out completely, and also require carefully worked prisms of fairly large dimensions. They are even more troublesome when applied to the case of liquids; as the liquid has to be contained in a hollow prism, built up of accurately plane parallel pieces of glass, which has frequently to be taken to pieces for thorough cleaning.

There is a large amount of refractometry, which has to be done in connection with various industries, for which the very highest accuracy is not demanded, and for which the time and trouble involved in such determinations would be unjustifiably expensive. For such purposes a number of instruments have been devised, most of which are largely empirical in their results, but which are very rapid in use, and require a minimum of material and labour for the preparation of the necessary specimens.

§ (12) ABBÉ'S AUTOCOLLIMATION SPECTROSCOPE.—This instrument employs a modification of the minimum deviation method in which the telescope also serves as collimator. In the upper half of the focal plane of the eyepiece is a slit. The light enters by a hole in the side of the tube and is reflected through the slit by a small 45° prism. In the lower half of the field is a sharp pointer. The prism employed for the test is cut with an angle of 30°. The prism table is rotated until the light refracted at the first face of the prism meets the second face normally, and is returned along its path to form an image of the slit in the focal plane of the telescope. When this image coincides with the pointer, the normality of the beam to the second face of the prism is exact.

The angle of refraction is then equal to the angle of the prism. The angle of incidence, i, is found by rotating the telescope until it is normal to the first face of the prism, and the refractive index is obtained from the formula

$$\mu=\frac{\sin i}{\sin \theta}.$$

The method is not very convenient for accurate determinations of the refractive index, since the value of θ must be determined; but it gives quite accurate values for the dispersion. It has the advantage for commercial work of requiring only half as much glass as a 60° prism, but has the disadvantage over some of the empirical refractometers, described later, of requiring two first-class surfaces instead of only one.

§ (13) PULFRICH REFRACTOMETER.[2]—This employs the *Wollaston* method discussed theoretically in §§ (2) to (5). The main feature of the instrument is a block, A, *Fig.* 20, with one polished face horizontal and another vertical. The substance whose index it is desired to determine is placed in contact with the horizontal surface and a convergent beam of monochromatic light, L, is directed along the interface. In *Fig.* 20 (*a*) the arrangement employed for liquids is illustrated. A circular glass tube, C, is cemented to the block with a cement or wax which is not soluble in the liquid to be examined, some of which is poured into the cell thus formed. A portion of the incident light is refracted from the liquid into the block, ultimately emerging from the vertical face into the air, and is received by a telescope. As in the case of the critical angle method for determining the index of the prism itself, a band of light is produced in the field of view which has a sharp edge corresponding to the direction of emergence of a ray incident at grazing incidence on the first face. The angle of emergence, e, which this ray makes with the normal to the vertical face of the block is determined. In order to set the telescope normal to the face of the

[1] Guild and Dale, *Trans. Optical Soc.*, 1920, xxii. 13.

[2] C. Pulfrich, *Zeits. Instrumentenk.*, 1888, viii. 47; and 1895, xv. 389.

block it is provided with an autocollimating device, consisting of a small 45° prism fixed behind a portion of the cross-lines. This prism can be illuminated by a lamp placed to the side of the instrument. When the telescope is approximately normal to the face

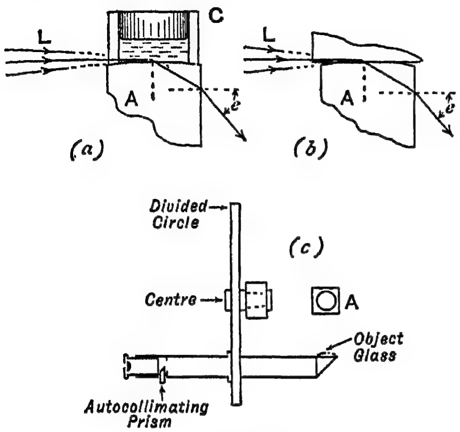


FIG. 20.

of the block a reflected image of the illuminated face of the little prism is seen at the side of the field opposite to the prism itself. It is crossed by images of the parts of the cross-lines which are in front of the prism. When the actual cross-lines are brought into coincidence with those images the axis of the telescope is normal to the face of the block.

To avoid having to place the head in inconvenient positions, the telescope is constructed as in *Fig.* 20 (*c*), with a right-angled prism behind the object glass, so that the main body of the telescope is always horizontal.

The circle is usually divided so as to read to minutes by means of a vernier. It is provided with a radius arm and micrometer tangent screw by means of which it may be moved through a range of 5° or 6°, for which readings can be taken on the drum of the tangent screw to 0·1 minute. Thus the actual values of e can be determined to about 1 minute, which, from the curves of *Fig.* 4, corresponds roughly to 1 unit in the fourth decimal place, whereas the differences of readings within the range of the tangent screw may be accurate to 0·1 minute, corresponding to about 1 unit in the fifth place. The quantities usually wanted in practice are the refractive index for some wave-length near the brightest part of the spectrum and the dispersion, *i.e.* the differences in refractive index for a series of lines spread throughout the spectrum. The lines almost universally used are C, D, F, and G′, the first, third, and fourth being obtained from the vacuum tube spectrum of hydrogen, and the second being the well-known sodium line. The procedure usually followed is to bring the tangent screw to the beginning of its range, and rotate the telescope so as to bring whichever of these lines corresponds to the smallest value of e, usually C, to the middle of the field. The circle is then clamped to the radius arm and the settings on the various bands made by means of the tangent screw. By noting the circle reading which corresponds to some reading on the tangent screw, and correcting for the reading which corresponds to the normal ($e=0$), the values of e for the various lines will be obtained correct to 1 minute, but mutually consistent to a much higher degree.

For determining the indices of solids, the arrangement of *Fig.* 20 (*b*) is employed. The specimen requires to have two polished faces, one of them two to three cm. long and accurately flat. The other, which should be approximately at right angles to the first, need only be a few millimetres deep, and its quality is immaterial provided it intersects the first surface in a clean sharp edge without trace of bevel.

In order to obtain optical contact, so that refraction from the specimen to the block may occur, it is necessary to interpose a film of liquid of higher refractive index than the specimen. If the film is parallel-sided—which can always be secured by examining the interference bands at the interface—the presence of the film does not affect the ultimate direction of the rays, which is the same as if the specimen were in optical contact with the block.

For this refractometer, in which the angle of the block A is 90°, equation (6), § (2), becomes

$$\mu_1^2 = \mu_2^2 - \sin^2 e.$$

For convenience in computation the instruments are usually supplied with tables giving the indices corresponding to values of e, increasing by intervals of 10 minutes, for each of the lines C, D, F, and G′. The values for intermediate angles are obtained by proportionate interpolation.

Unfortunately such tables as issued are rarely sufficiently accurate for any but the roughest approximation. The actual angle of the block usually departs appreciably from 90°; and the values adopted for the indices of the block are those determined for a specimen of the melting used for a number of blocks, and may differ considerably from those actually possessed by any one of them.

The angle should always be measured, and the true indices of the block deduced from careful determinations with a substance whose indices have been determined by an accurate absolute method; or, if this is not available, with a specimen of quartz, for which the

values are reasonably well established. Having obtained this information a new set of tables can be computed; but it is much simpler to employ the erroneous tables and apply corrections.

For instance, suppose the angle of the prism is $90° - \delta\theta$, where $\delta\theta$ is a small quantity.

$$\mu_1 = \bar{\mu}_1 - \frac{\delta\theta}{T_\theta}$$

$$= \bar{\mu}_1 - \frac{\delta\theta \sqrt{\mu_2^2 - \mu_1^2}}{2{\cdot}06} \text{ (from 13),}$$

where $\bar{\mu}_1$ is the value given by the tables. The tabulated value, $\bar{\mu}_1$, has to be diminished to give the true value if the angle of the block is less than 90°, and increased if the angle is greater than 90°. The correction can be calculated for a series of values of $\bar{\mu}_1$, say 1·5, 1·55, 1·6, 1·65, etc. It will not be necessary to tabulate it for smaller intervals unless the error in angle is unusually large.

A similar small correction can be applied to the tables for errors in the assumed values of μ_2. This of course has to be done for each wave-length. Let $\bar{\mu}_2$ be the assumed value used in computing the tables, and μ_2 the true value. Also let $\bar{\mu}_1$ be the value of the index under test as given by the tables, and μ_1 the true value.

$$\mu_1 = \bar{\mu}_1 - \frac{\partial\mu_1}{\partial\mu_2}(\bar{\mu}_2 - \mu_2)$$

$$= \bar{\mu}_1 - \frac{\mu_2}{\mu_1}(\bar{\mu}_2 - \mu_2).$$

This correction can also be tabulated for a series of values of $\bar{\mu}_1$ for each wave-length. If there is also an error in the angle of the block, the two corrections, being small, can be obtained separately as above and added to give a single correction.

The total correction for values of $\bar{\mu}_1$ increasing by intervals of ·01 can be got out in a comparatively short time. The corrections should be computed to the sixth decimal place to prevent accumulations of small errors mounting up to 1 in the fifth place.[1]

§ (14) THE ABBÉ REFRACTOMETER. — This instrument is of lower accuracy than the Pulfrich refractometer and gives results which are only reliable to a few units in the fourth decimal place.

The essential parts of the instrument are shown diagrammatically in *Fig.* 21, the arrangement of parts being that followed in the instrument made by Bellingham and Stanley of London, which has various advantages over the Zeiss model. P_1 and P_2 are two prisms of dense glass. Surfaces 1, 3, and 4 are polished; surface 2 is matt. The prism P_1 is hinged at H, so that it can be swung away from P_2 and removed altogether if desired.

In determining the index of a liquid, a drop of the latter is placed on surface 2, which is then closed up into contact with surface 3, the liquid being squeezed out into a thin film. Light from a suitable source is directed towards the prism system by a mirror M. It

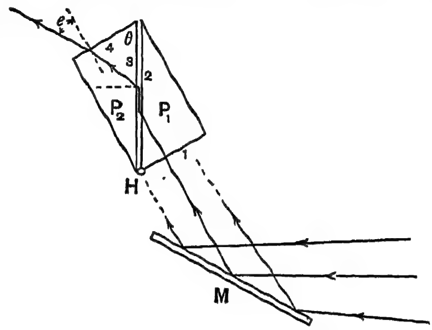


FIG. 21.

strikes the matt surface 2, and is scattered into the liquid film and the prism P_2. Since no ray can enter P_2 with a greater angle of refraction than that of the ray at grazing incidence, the emergent rays when collected by a telescope will all converge to points on one side of a line in the focal plane. The field will therefore be divided into dark and light portions, the edge of the bright portion corresponding to the value of e for the critical rays, as in the similar cases previously discussed.

The rotation of the telescope is registered on a graduated arc which reads refractive index directly. In practice the instrument is used with white light. To compensate for the dispersion of the system, and at the same time get a rough measure of the dispersion of the substance under test, the ingenious device is adopted of introducing a reverse dispersion by means of two direct vision prisms. These are mounted one above the other in front of the object glass of the telescope, and are geared so that by turning a milled head they may be rotated in opposite directions. They are so oriented to start with that their planes of dispersion are parallel to each other and to the principal plane of the prism P_2, their dispersions being additive. If k is the angular dispersion of one of the prisms between the C and F lines, the total dispersion in this position is $2k$. If now the prisms are rotated at equal rates in opposite directions, the resultant dispersion is still parallel to the principal plane of P_2, but is diminished. For any orientation, ϕ, from the initial position the resultant dispersion is $2k\cos\phi$. When ϕ is 90° the dispersion due to the prisms is zero, and when ϕ exceeds 90° it changes sign, and increases to $2k$ in the reverse direction.

In using the instrument the critical edge as

[1] For a discussion of the possibilities of the Pulfrich Refractometer and the limitations of existing patterns of the instrument, see Guild, *Proc. Phys. Soc.* xxx. part 111, p. 157.

first observed will probably be spread out into a short spectrum. The compensating prisms are rotated until the colour disappears and a sharp achromatised edge is obtained. The cross-lines are set on this edge; the reading on the arc then gives the index for sodium light. The orientation of the compensating prisms gives, by reference to a set of tables supplied with the instrument, a rough value of the dispersion, from C to F, of the substance under test.

To use the Abbé refractometer to determine the indices of solids, the prism P is removed, and a specimen similar to that employed with the Pulfrich refractometer placed against face 3, optical contact being obtained by a film of liquid, as described in connection with the latter instrument.

§ (15) ABBÉ'S CRYSTAL REFRACTOMETER.[1]—This instrument is designed for the convenient examination of crystals in different orientations. The block is in this case hemispherical, *Fig.* 22, and is fitted with a cell to contain

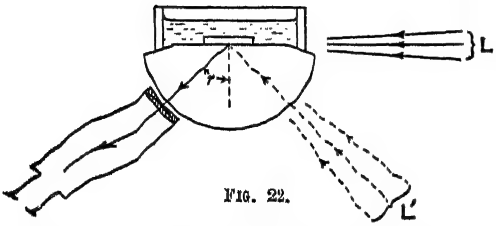


FIG. 22.

liquids. The crystal specimen is placed on the surface, immersed in a liquid of higher refractive index. The axis of rotation of the telescope passes through the centre of the hemisphere, so that the optic axis is always normal to the latter. The front surface of the object glass is close to the hemisphere and parallel to it, thereby neutralising the effect of the curvature of the block on the focus of the emergent beam. There is evidently no deviation of the beam on leaving the block, so that when the telescope is set on the critical edge it measures r, the angle of refraction of the critical ray into the block. The index of the specimen, $\mu_1 = \mu_2 \sin r$.

It is possible to rotate the hemisphere about a vertical axis, thereby altering the orientation of the crystal with respect to the direction of the light. The illumination may either be supplied at grazing incidence, as in the Pulfrich refractometer, or the beam may be incident from within the glass as shown dotted at L′. In the latter case the critical edge divides the field into regions of total and partial reflection, the critical direction being the same as for grazing incidence. The arrangement is advantageous with specimens whose edges are irregular or not approximately perpendicular to the face in contact with the block of the instrument; but the edge is not so sharp nor the contrast so good as with grazing incidence. In order to pass readily from one method of illumination to the other, the source is mounted in line with the axis of rotation of the telescope, the light being reflected in the direction L or L′ by a mirror mounted on an arm which rotates about the same axis. Thus, by swinging the mirror, the light can quickly be adjusted for either type of illumination.

[1] See S. Czapski, *Zeits. f. Instrumentk.*, 1890, x. 246, and Feussner, *Zeits. f. Instrumentk.*, 1894, xiv. 87.

§ (16) THE DIPPING REFRACTOMETER.—This instrument is designed for the refractometry of liquids available in fairly large quantities. It is only suitable for use over a small range of index and is therefore mainly used for special purposes, such as the estimation of alcohol, wines and beers, sea and mineral waters, etc. The principle of the apparatus will be clear from *Fig.* 23. P is a cylinder of glass on which a plane surface inclined at about 30° to the axis has been ground and polished. The end of the instrument is immersed in the liquid, which is contained in a vessel with a glass side so that light reflected from a diffusing reflector, R, may enter the vessel. The critical edge is viewed in the focal plane of the eyepiece, in which there is also a scale divided to read refractive indices directly. A is a direct vision prism which compensates the

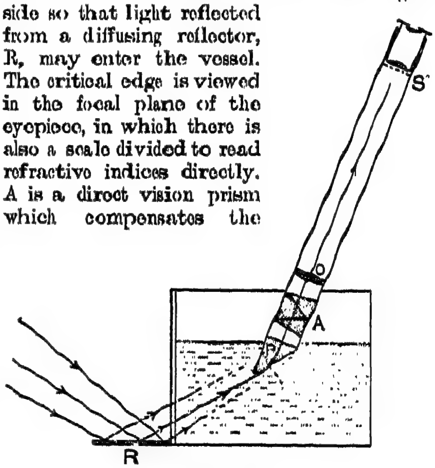


FIG. 23.

dispersion and ensures a sharp achromatic edge. The instrument is usually supplied with several nose-pieces of different glass to render it suitable for different ranges. The most usual range is 1·325 to 1·365.

IV. EFFECTS OF TEMPERATURE AND PRESSURE

§ (17). In the preceding paragraphs we have concerned ourselves with the methods of determining refractive indices as accurately as possible, *under the circumstances which prevail during the determination*. For the results to

be of value it is necessary to state them in such a form that they can be compared with those made under other conditions. In the first place, the quantity measured by all the methods employed in practice is the *relative* index, *i.e.* the absolute index of the substance with respect to vacuum divided by the refractive index of air. If $\bar{\mu}$ is the absolute index and μ the observed index, $\bar{\mu}=\mu n$, where n is the refractive index of the air during the observation. Both $\bar{\mu}$ and n are affected by temperature, and n is also affected by pressure. An observed value of μ is therefore only correct for the particular temperature and pressure for which the determination was made.

As regards the variation of $\bar{\mu}$, this may be taken as linear over the usual range of laboratory temperatures, so that $\bar{\mu}_t=\bar{\mu}_{t'}+\mu a(t-t')$, where a is the temperature coefficient of absolute index, and is taken as positive when the index increases with temperature.

The variation of n with temperature and pressure follows Gladstone and Dale's law that $n-1$ is proportional to the density of the gas. Thus

$$n_t^p-1=\left(n_0^{760}-1\right)\frac{p}{760}\cdot\frac{273}{273+t}$$

$$=\beta\left(1+\frac{\delta}{760}\right)\left(1-\frac{t}{273}\right)=\beta+\beta\frac{\delta}{760}-\beta\frac{t}{273},$$

where $1+\beta$ is the refractive index of air at 0° C. and 760 mm. pressure, and δ is the *excess* of pressure above 760 millimetres.

Let μ_t^δ be the observed relative index at t° C. and $760+\delta$ mm. pressure, then

$$\bar{\mu}_t=\mu_t^\delta n_t^\delta=\mu_t^\delta\,(1+n_t^\delta-1)$$

$$=\mu_t^\delta+\mu\beta+\mu\beta\,\frac{\delta}{760}-\mu\beta\frac{t}{273}. \quad \text{(i.)}$$

This gives $\bar{\mu}_t$ in terms of the observed relative index. Similarly, for observations at any other temperature and pressure,

$$\bar{\mu}_{t'}=\mu_{t'}^{\delta'}+\mu\beta+\mu\beta\,\frac{\delta'}{760}-\mu\beta\,\frac{t'}{273}. \quad \text{(ii.)}$$

Subtracting (ii.) from (i.) and rearranging, we get

$$\mu_{t'}^{\delta'}=\mu_t^\delta+\mu\beta\frac{\delta-\delta'}{760}-\mu(t-t')\left(a+\frac{\beta}{273}\right). \quad \text{(iii.)}$$

Equation (iii.) gives the relative index at any temperature and pressure in terms of the relative index at the temperature and pressure of the observation. It is evident that the temperature coefficient of the *relative* index is $a+\beta/273$.

These relationships enable us to express a result either as *absolute* index at a standard temperature, or as *relative* index at a standard temperature and pressure. The most useful in practice is the latter, and, where possible, results should be expressed for a temperature of 17° C. and a pressure of 760 mm.

The values of β can be obtained from the results of Kayser and Runge,[1] Scheel,[2] and others. Kayser and Runge's formula for *damp* air at 0° C. and 760 mm. pressure is

$$\beta=10^{-7}\left(2878\cdot7+13\cdot16\frac{1}{\lambda^2}+0\cdot316\frac{1}{\lambda^4}\right),$$

where λ is the wave-length of the light. For dry air 3×10^{-7} should be added. This difference is negligible for our present purpose. The following table gives the indices for dry air for various spectrum lines:

Designation in Solar Spectrum.	Approximate λ.	β.
A	7590 AU.	·0002905
B	6870 „	·0002911
C	6560 „	2914
D	5890 „	2922
E	5270 „	2933
F	4860 „	2943
G	4310 „	2962
H	3970 „	2978

From these figures the reduction of the results to standard pressure can always be performed; and, if a is known, the reduction from one temperature to another. The coefficients both of absolute and relative index are given for a typical series of optical glasses in Hovestadt's "Jenaer Glas." For the ordinary flint glasses a varies fairly regularly with μ, and the value for any glass of this type can be found sufficiently accurately for correction purposes by interpolation. For the more complex glasses, however, the data are insufficient for interpolation, except for glasses which are very near one of the type glasses, and it may be necessary to determine a. In doing this it is safest to avoid attempts to heat the prism. It is very difficult to obtain reliable results in this way. In most laboratories there is some room or rooms of which the temperature can be varied over a range of at least 10° C., so that measurements can be made, with the whole apparatus at a uniform temperature, over a sufficient range to give the necessary accuracy in a for correcting over a smaller range.

V. Special Varieties of Spectroscopes

§ (18). The type of spectrometer discussed in § (6) is really in principle the simplest and earliest form of spectroscope, elaborated in mechanical detail so as to enable the measure-

[1] *Abhandl. d. Berl. Akad. Wiss.*, 1893.
[2] *Berl. Verh. d. Physik. Ges.*, 1907, ix.

ments involved in refractometry and goniometry to be made with the utmost accuracy. Except for these specific purposes, however, such an instrument would rarely be used. It would find no place, for instance, in the equipment of a modern spectroscopic laboratory, where the primary object of the work is the identification and accurate measurement of wave-lengths. Such determinations are always made by interpolation between adjacent lines of known wave-length, and the spectroscopist is never concerned with any but quite small angular measurements. This being so, the instrumental requirements are somewhat different. Not only so, but the development of photographic methods has practically ousted the spectro*scope* as an instrument of spectroscopic research.

Thus the beautifully designed instruments, with long trains of prisms intended to give very high resolving power, which were formerly features of spectroscopic laboratories, are now rarely if ever employed; they are replaced by spectrographs, which not only are capable of greater accuracy, but have the inestimable advantage of giving permanent records of the observations. The visual spectroscope is only used for the identification of simple known spectra, or for other work for which great accuracy is not required. The special needs of the spectroscopic laboratory are beyond the purview of this article, which deals primarily with the requirements of the "applied" rather than the "pure" physicist; but a certain amount of spectroscopic equipment, suitable for general work, is essential in an optical laboratory.

§ (19) The Direct Vision Spectroscope.—This instrument, which is suitable for the examination of light sources to ascertain the general nature of their spectra, has all its parts mounted in one straight tube, and is therefore easily directed to the light when simply held to the eye. Its essential element is a compound prism, P (*Fig.* 24), which consists of a

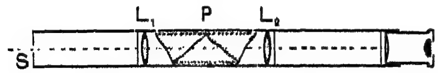


Fig. 24.

prism of dense flint glass cemented with Canada balsam between two prisms of crown glass. The refractive index of the denser glass will be about 15 per cent greater than that of the other, but its dispersion may be twice as great. Consequently, if the prisms be cut to such angles that the total deviation for, say, the sodium light is zero, rays at the blue end of the spectrum will be deviated towards the base of the dense prism, and those at the red end will be deviated in the opposite direction, on account of the preponderating dispersion of the flint glass.

The collimating and telescope lenses, L_1 and L_2, have usually a focal length of 2 to 3 inches. In some instruments the slit, *s*, is fixed, and the positions of lines in the spectrum are read on a scale in the focal plane of the eyepiece. This is only convenient for instruments of small dispersion. If the dispersion is large the whole spectrum cannot be seen in the field of view at once. In a convenient form of instrument made by Hilger the slit can be moved across the end of the tube by a tangent screw. By this means different parts of the spectrum can be brought to the centre of the field of view, which is marked by a pointer.

Larger and more ambitious types of direct-vision spectroscope can be obtained; but for anything except the roughest of purposes it is better to employ—

§ (20) The Hilger Constant Deviation Spectroscope.—This is an exceptionally convenient instrument for general laboratory use. *Fig.* 25 (*a*) is a diagrammatic plan. The collimator C and telescope T are permanently fixed at right angles to each other. The prism P is clamped to a table which may be rotated through a small angle by means of a tangent screw. The nut in which the tangent screw works is attached to the fixed part of the stand, and its movement is registered by a drum D with spiral scale. The special feature is the prism, which can be regarded as built up of three prisms arranged as shown in *Fig.* 25 (*b*). If a ray of light is incident at such an

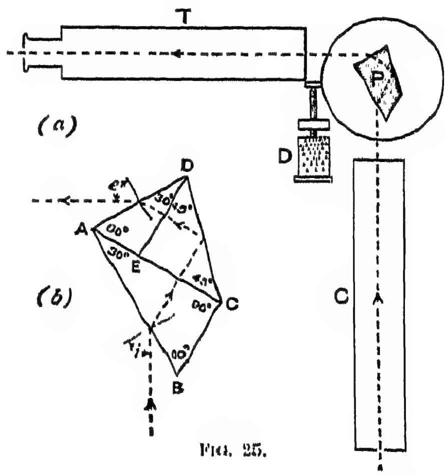


Fig. 25.

angle on AB that it is refracted parallel to BC it will be parallel to EA after internal reflection at the hypotenuse face of the 45° prism CED. It is therefore incident on AD with an angle of incidence equal to the angle

of refraction at AB, and so emerges from the prism at an angle, e, which is equal to the angle of incidence i. Since $D\hat{A}B=90°$ and $i=e$, the incident and emergent rays are at right angles. But we may regard ACB and ADE as two halves of a 60° prism, so that a ray which travels parallel to BC and EA suffers minimum refraction.[1] This applies to any ray which traverses the prism and emerges at 90° to its initial direction. Clearly the shorter the wave-length, and, consequently, the greater the refractive index of the prism, the greater the angle i will have to be in order that the refracted ray should be parallel to BC. Thus as the prism is rotated by the tangent screw, thereby altering the angle of incidence of the beam from the collimator, the spectrum will pass across the field of view, each wave-length being at minimum deviation, in the restricted sense indicated in the footnote, at the moment when it occupies the centre of the field.

The advantages of this instrument over an ordinary 60° prism spectroscope are, firstly, that both telescope and collimator being fixed extreme rigidity and permanence of adjustment are obtained; and, secondly, no readjustments of the prism are necessary in order to get minimum deviation for all measurements. The drum is graduated to read wave-lengths directly; and all that is necessary is to adjust the cross-lines of the eyepiece occasionally so as to make the instrument read correctly for some known wave-length.

§ (21) Monochromatic Illuminators.—It is frequently necessary to obtain monochromatic light, either from a source giving a continuous spectrum, or by isolating a particular line in a line spectrum. Any spectroscope can be used for this. It is only necessary to replace the eyepiece by a second slit in the position normally occupied by the cross-lines. If the collimator slit is illuminated by white light, or light containing more than one spectrum line, the light which passes through the second slit can only contain the wave-length of that part of the spectrum which falls on it. Either the direct vision spectroscope described above, or the constant deviation spectroscope, make excellent monochromatic illuminators. The former is particularly suitable for separating the green line of the mercury spectrum from the other lines. If the slit widths are chosen as wide as possible without including any of the nearest yellow line, the resulting illumination is very bright, and suitable for such purposes as polarimetry where an intense light is required. For use with white light greater dispersion than that of the direct-vision instrument is required, and the constant-deviation spectroscope is better. Hilger makes a special form of the instrument for use as a monochromatic illuminator. The telescopes are shorter and of larger aperture, thus securing a greater amount of light, and both slits open symmetrically. On account of the fixed direction of the emergent beam, the Hilger instrument is specially suitable for purposes in which it may be necessary to vary the wave-length during an experiment. The importance of being able to do this will be realised in reading the article on Immersion Refractometry.

There is one precaution which must be borne in mind in using monochromatic illuminators for work, such as polarimetry, in which complete exclusion of other wave-lengths is essential. There is always a proportion of the light which passes through any spectroscopic system scattered irregularly at the surfaces of the lenses and prisms, and even within the material. This is due to imperfect polishing, minute air-bubbles, and to any dust particles that may be present. Some of the scattered light, which contains all wave-lengths present in the light from the source, passes through the slit in addition to the rays which are properly focussed there. Thus, in separating the green line of mercury, if the light from the monochromatic illuminator is examined by another spectroscope it will be found to contain an appreciable quantity of the yellow and blue radiations. For most purposes this is immaterial, but for polarimetry it is fatal, on account of the rapid variation of optical rotations with wave-length. It is necessary in this case to pass the light from the first illuminator through a second, which reduces the scattered residuum to a negligible amount.

J. G.

[1] The total deviation, which is due not only to refraction but to reflection at DC, is never a minimum. It is only the portion due to refraction that has a minimum value; but under these conditions the magnification is unity, as in the case of a 60° prism at minimum deviation. It is only in this restricted sense that the term minimum deviation can be employed in connection with this instrument.

SPECTROSCOPY, MODERN

§ (1) Apparatus and Methods of Observation.—In the classical researches of Bunsen and Kirchoff, one of the chief functions of the spectroscope appeared to lie in its use as a tool in chemical analysis and in the discovery of new elements, and indeed it may be said that many of the chemical elements would probably not have been detected or isolated without the help of the spectroscope. In addition to the spectroscopic detection of elements, by examination of the radiations which they emit when in the state of luminous gases, the spectroscope has proved of signal service in the identification of compounds, particularly in organic chemistry, by observation of the radiations which they absorb, either in a pure state or

when dissolved in appropriate solvents. At the present time, however, the main uses of the spectroscope are in its application to physical problems, in which it has proved one of the most powerful methods of attack in many widely differing fields of investigation. In its simplest form the spectroscope consists essentially of a slit on which the image of a luminous source is projected, the rays which traverse the slit being made parallel by means of a lens, known as the collimating lens, and after passing through the dispersing system, consisting of one or more prisms, being brought to a focus by means of a telescope lens, the spectrum being either observed visually with an eyepiece or recorded permanently on a photographic plate. This simple system forms the basis on which the modern spectrometer has been developed, and can be used in this form for preliminary visual observation and approximate measurement of wave-lengths. In recent years, however, visual observations have been almost entirely superseded by photographic methods, for the latter have numerous advantages both in regard to accuracy of measurement, the range of wave-lengths which can be recorded, and owing to the fact that a considerable portion of the spectrum can be photographed at the same time on the plate, thus obviating the instrumental displacements which are liable to occur when consecutive measurements are made. Spectrographs were formerly designed with trains of prisms of relatively small aperture and lenses of short focal length, but in recent years there has been a tendency towards single prism instruments of large aperture, the required dispersion being obtained by the use of lenses of greater focal length. In the design of all forms of spectrograph, rigidity of all the parts is an essential feature, since exposures often extending over many hours have to be made, and any displacement of the apparatus would give rise to serious errors in measurement or loss of definition. In designing a spectrograph for a particular purpose it is generally necessary to arrive at some compromise between dispersion and light-gathering power, the former attribute being essential for the accurate determination of wave-lengths and the latter for recording faint spectra which would require impracticably long exposures with an instrument of low light-gathering power. For the accurate determination of wave-lengths either concave gratings or plane gratings mounted with lenses of long focus are to be preferred, though good results can be obtained with prism instruments of adequate dispersion. For the photography of faint spectra it is often essential to use prisms of large aperture and lenses of relatively short focal length with a consequent sacrifice of dispersion. For the visible spectrum and the ultra-violet spectrum down to about 3600 Å. the lenses and prisms are usually made of glass, but for the more refrangible rays down to 1800 Å. both lenses and prisms are made of quartz, since glass is opaque to these radiations. In the latter case the quartz prisms are made up of two halves of right- and left-handed quartz respectively, and the collimating and camera lens are also made of right- and left-hand quartz in order to obviate the doubling of the image which would otherwise occur. An instrument which is now widely used and which has many advantages is known as the Littrow spectrograph. This mounting, in which the principle of autocollimation is used, is applicable both to gratings and prisms and is essentially similar to the Abbé spectrometer. In the determination of wave-lengths a spectrum containing a sufficient number of lines of known wave-lengths is photographed in juxtaposition to or superposed on the spectrum to be measured. The wave-lengths of the unknown lines are determined by measurement of their positions with respect to the standard lines and subsequent reduction by appropriate formulae. In all accurate work the utmost care must be taken to ensure correct superposition of the spectra and to guard against instrumental displacement between successive exposures. A great deal of attention has been paid to the photographic plates used for spectroscopic purposes, and whereas ordinary plates are insensitive to wave-lengths greater than about 5000 Å. it is now possible to obtain plates which have been bathed in appropriate dyes which render them sensitive as far as the extreme visible red, and by the use of dicyanine bathed (1) plates it has recently been found possible to extend the sensibility of the photographic plate to about 9600 Å. In the regions of shorter wave-length the ordinary plate is sensitive to about 2000 Å., but beyond this limit its sensibility falls off rapidly owing to the absorption of the more refrangible rays by the gelatine of the plate. Plates are now made on a commercial scale in which this difficulty is overcome by following the special methods of preparation devised by Schumann (2), in which the gelatine is reduced to a minimum, and such plates have recently been used successfully by Millikan down to a wave-length of 360 Å. It should be mentioned that a number of special methods, notably that of Abney (3), have been devised for the purpose of extending the sensibility of photographic plates into the infra-red, but the technique of these methods is exceedingly difficult and laborious and they have not come into general use. In spectrographs in which lenses of single material (not achromatic lenses) are used, it is necessary to incline the plate, so that the rays do not fall on the plate at an angle perpendicular to its surface, and in such cases

it is essential that backed plates should be used to eliminate the reflection from the back of the plate, though it is desirable to use backed plates in all cases in which the most critical definition is required. For the photography of the ultra-violet regions below 1800 Å. it is necessary to use specially designed spectrographs from which the air can be exhausted or which are filled with pure helium, since air absorbs these radiations strongly. For the examination of the infrared regions of the spectrum which are beyond the range of sensibility of photographic plates special spectrometers are used. In these instruments glass must be eliminated and either gratings, or prisms of rock-salt or sylvin, are used. Lenses are usually replaced in these instruments by concave reflectors of metal, and a second slit is placed at the focus of the telescope reflector. Radiations which traverse this second slit are detected and their intensity is measured by means of a thermopile or a bolometer which is fixed behind the slit.

§ (2) Classification of Spectra.—Luminous spectra can be divided into two classes, namely continuous spectra, which result from the radiation of heated solids, and discontinuous spectra, which are in general peculiar to luminous gases, though in some cases continuous spectra can be observed from luminous gases, and solids may under certain conditions give rise to spectra which cannot be described as continuous. Discontinuous spectra may be subdivided into line and band spectra, the former consisting of lines distributed at intervals through the spectrum in a manner in which certain mathematical relations can frequently be found, though the existence of such relations is not immediately apparent; band spectra usually consist of complex groups of lines arranged in a manner in which some regularity is obvious, these groups repeating themselves at intervals in the spectrum. The regularities in these two classes of spectra are discussed in a later section.

§ (3) The Production of a Spectrum. (i.) *Flame Spectra.*—One of the most striking phenomena in spectroscopy is to be found in the variation in the spectrum of an element under different conditions of excitation. The spectrum of substances introduced into the Bunsen flame consists generally of a certain number of lines due to the element together with bands due to compounds of the element with oxygen or with other elements with which it has been introduced into the flame. Thus in a spectrum of calcium chloride introduced into the Bunsen flame we have, in addition to the characteristic flame lines of the element, bands due to calcium oxide and other bands due to calcium chloride, the spectrum of calcium bromide being thus similar as regards the lines and oxide bands but showing bands peculiar to calcium bromide in place of those due to calcium chloride. The efficacy of flames, as a method of producing spectra, can be greatly enhanced by introducing the substances in a finely divided state, most conveniently in the form of a spray (4), but the method is in general limited to the more volatile elements.

(ii.) *Arc Spectra.* — The spectra of less volatile elements can be conveniently produced by burning an electric arc between poles of a metal under investigation or between carbon poles impregnated or filled with the substance in the case of non-metallic bodies. When the arc burns between metallic poles the spectrum is found to consist of lines due to the element and sometimes also of bands due to the oxide; but in the case of substances introduced into the carbon arc there are a vast number of lines due to impurities in the carbon poles and bands which are attributed to cyanogen and other carbon compounds. The number of lines exhibited by different elements under these conditions varies widely, as also does the quantity of an element necessary for the appearance of its spectrum. The distribution of intensity amongst the lines in the spectrum of a given element depends also on the density of the radiating vapour, and it is found that the last lines to disappear from the spectrum when the quantity of a substance in the arc is gradually reduced are not necessarily the lines which are strongest when the density of vapour in the arc is greater (5). The recognition of the last lines to disappear under these conditions is, of course, a matter of importance for the detection of small quantities of a substance which is present as an impurity. The electric arc, burning between iron poles, is a source of radiation of great importance in spectroscopy, since its spectrum contains the lines selected as the international secondary standards of wave-length, and a special form of the iron arc, due to Pfund (6), has been chosen as the standard iron arc for this purpose on account of the fineness of the iron lines which it yields. Arc spectra generally contain, in addition to the "arc lines" proper, the lines characteristic of the flame spectrum and also a certain number of lines which belong more appropriately to the spark spectrum. A number of interesting changes occur when the arc is burnt in vacuo. Under these conditions the spark lines, which are relatively feeble in air, are greatly enhanced, and the oxide bands disappear and are often replaced by bands usually attributed to hydrides. The arc in vacuo thus represents a transition stage between the arc in air and the spark spectrum, and it has the advantage that the spark lines

are exceedingly narrow and therefore capable of measurement with a high degree of precision, which is unobtainable in spectra of condensed spark discharges, in which the spark lines, though relatively stronger in so far as their energy content is concerned, are so diffuse as to be unsuitable for measurement.

(iii.) *Spark Spectra.*—A striking change in the spectra of most substances is found when the arc spectrum is compared with the spectrum obtained when high potential discharges from an induction coil or transformer are allowed to pass between poles of the substance under investigation. With discharges obtained without the use of a condenser the spectrum consists mainly of bands due to air with feeble metallic lines in the neighbourhood of the poles, but when a condensed discharge is employed the spark spectra of the poles are obtained together with the characteristic spark lines due to the nitrogen and oxygen of the air. Whilst the latter lines extend uniformly across the spark gap the lines due to the poles are strongest in the immediate neighbourhood of the poles and extend to varying distances across the spark gap (7). Both arc and spark lines are present, the latter being greatly enhanced in intensity, and being in many cases found only in the immediate neighbourhood of the poles, whilst the arc lines extend to a considerable distance between the poles or right across the gap. The lines which extend furthest across the gap are precisely those lines which are the last lines to disappear when the amount of the substance in an electric arc is gradually reduced. The spark lines themselves extend to varying distances across the gap and can thus be classified into different groups. A remarkable change is observed when the period of the condenser is increased by the introduction of a self-induction coil in the circuit (8). Under these conditions the air lines disappear from the spectrum, which in other respects approximates closely to that of the electric arc, only the "longest" of the spark lines being visible in the immediate neighbourhood of the poles. The temperature of the poles also exerts an influence on the spectrum, the effect of cooling being in general to enhance relatively the intensity of the spark lines.

(iv.) *Tube Spectra.*—Gases can best be excited to luminosity by the use of vacuum tubes. Such tubes generally consist of two glass bulbs connected by a length of capillary tubing, each bulb being provided with an electrode which is usually made of aluminium. The tubes contain the gases to be investigated at relatively low pressures, generally in the neighbourhood of a few mm. of mercury, and can be excited by means of an induction coil or transformer. The preparation and filling of these tubes require a great deal of care, as the spectrum of the substance which it is desired to examine is often masked by impurities, notably compounds of carbon, with which the glass tube and the electrodes are frequently contaminated. In preparing these tubes they are usually exhausted to a very high vacuum, and there are several methods by which the desired gas can then be admitted into the tube in a sufficiently pure state, but special methods are usually required for different gases. As in the case of the spectra of spark discharges between metallic poles, the spectra of gases in vacuum tubes can be profoundly modified by the introduction of a condenser and spark gap into the electrical circuit. Nitrogen, for example, shows with an uncondensed discharge two band spectra, one of which is known as the negative band spectrum, and which appears in the neighbourhood of the cathode, whilst the other, the positive band spectrum, is brightest in the capillary portions of the tube. When the condenser and spark gap are introduced into the circuit the two band spectra disappear and are replaced by a bright line spectrum, which is the spark spectrum of nitrogen. This spectrum is itself capable of further modification, for when more powerful spark dischagres are employed certain of the lines are relatively enhanced. With these powerful condensed discharges, it is usual to find lines due to the constituents of the glass walls of the capillary. A very striking example of the change due to the introduction of the condenser and spark gap is afforded by the rare gas argon, which with the uncondensed discharge glows with a red light, which changes to a bright blue when the condensed discharge is used. In addition to spectra due to pure substances, it is common to find in vacuum tubes bands due to compounds. A large number of different types of vacuum tube have been designed for different purposes. For the investigation of the ultra-violet it is necessary to use tubes of fused silica, or glass tubes provided with a quartz window. For the investigation of some substances, *e.g.* metallic cadmium, which are not volatile at ordinary temperatures, it is possible to obtain satisfactory results by raising the temperature of the vacuum tube. In tubes containing mixtures of gases it frequently occurs that the gas which is present in smaller quantity dominates the spectrum, though it is often possible to reverse this state of affairs by a change in the conditions of electrical excitation. Thus in a mixture of helium and argon a small quantity of argon is sufficient to mask the helium spectrum when an uncondensed discharge is used, although the helium lines appear brightly when a condensed discharge is employed. Some gases are decomposed by the passage of the electric discharge, and

in these cases it is necessary to arrange for a continuous flow of the gas through the tube so that the products of decomposition are carried away. Changes in the spectrum of gases in vacuum tubes can also be brought about by changes in the pressure in the tube, and by the presence of impurities. Some of these changes are very striking and can be observed without difficulty, whilst others are discovered only when photometric measurements are made of the relative intensities of the lines.

(v.) *Other Methods of Production.*—Interesting results have been obtained by examining the spectra of substances at very high temperatures by enclosing them in graphite tubes heated by the passage of a very powerful electric current (9). In such cases the luminosity is probably due to temperature alone, but it resembles in general the spectra from electric arcs, although at the highest temperatures available a certain number of spark lines can be observed. In recent years attention has been paid to the spectra of fluorescent substances (10). In the case of solids and liquids, the spectrum usually consists of broad and somewhat indefinite bands, but in the case of fluorescent vapours the spectra are discontinuous and often exceedingly complex, the spectrum depending on the nature of the exciting light.

§ (4) Estimates of Intensity.—In recording a spectrum it is usual to give, in addition to the wave-lengths of the lines, an estimate of their intensities and an indication of any peculiar characteristics which they show. Thus, in a given spectrum some lines may appear perfectly sharp, whilst others are diffuse, either uniformly or in the direction of longer or shorter wave-length. It is usual to express intensities on an arbitrary scale from 0 to 10, the strongest lines being designated 10, whilst those which are just visible are given as intensity 0. This procedure, whether by direct visual observations or from photographic plates, is very inaccurate and is subject to considerable personal error on the part of the observer. In the infra-red regions of the spectrum where the thermopile or bolometer are used, the intensities can be quantitatively measured, and methods have been introduced recently by which quantitative measurements can also be made photographically in the visible and ultra-violet portions of the spectrum (11).

§ (5) Doppler's Principle.—There are a number of phenomena in spectroscopy which depend on the application of Doppler's principle. According to this principle a small change in the wave-length of a line is found when the observer or the source of radiation is in motion. If λ_0 is the wave-length observed when both the observer and the source are at rest, and if a is the velocity of the observer and b that of the source in the line of sight, V being the velocity of light, then λ, the observed wave-length when the observer or source is in motion, is given by the equation

$$\lambda = \lambda_0 \left\{ \frac{V \mp b}{V \pm a} \right\},$$

the upper sign being used in the case of approach.

Thus, if the displacement of a line is measured, the velocity of the moving source relative to the observer can be determined. This has been applied to the determination of the velocity of the positive rays in hydrogen and other gases (12). In such experiments, in addition to the displaced line, there is a line in the undisplaced position, the latter being due to particles made luminous in the path of the positive rays, whilst the former is due to the radiation of the moving particles themselves. According to the kinetic theory of gases the particles of a gas are in motion, and in consequence part of the radiation in a luminous gas will be from particles which are approaching the observer, and some from particles which are receding. The result of this is that spectrum lines are never infinitely narrow, but are broadened according to a law which depends on the distribution of velocities in the gas. These velocities are proportional to the square root of the absolute temperature, and inversely proportional to the square root of the masses of the radiating particles. According to the kinetic theory the distribution of intensity in a line which is broadened in this way follows a law of the form $y = e^{-kx^2}$. The width of a spectrum line is thus mathematically infinite, but it is usual to define it by the "half-width," which is the value of x when $y = \frac{1}{2}$. Lord Rayleigh (13) has shown that if $\delta\lambda$ is the "half-width" of a line of wave-length λ from a source at T° absolute, the masses of the radiating particles, in terms of that of the hydrogen atom, being m,

$$\frac{\delta\lambda}{\lambda} = 3{\cdot}57 \times 10^{-7} \sqrt{\frac{T}{m}}.$$

This affords a means of determining the masses of the radiating particles in luminous gases if the temperature is known, or the relative masses independently of the temperature in the case of mixtures of gases. By measuring the "half-widths" of the lines of the rare gases from vacuum tubes at ordinary temperatures, and when immersed in liquid air, it has been possible to show that the temperature of the radiating particles in a discharge tube is not appreciably greater than that of the walls of the tube, and that the measured "half-widths" of the lines are

in close agreement with the values deduced on theoretical grounds (14).

§ (6) Distribution of Intensity.—It has been found possible to measure the actual distribution of intensity in the lines, and the results have been found to be in accordance with theory. Investigations of this kind require the use of spectroscopic apparatus of very high resolving power, far exceeding that of any prism instruments and barely within the capacity of the largest gratings. The instruments usually employed for investigations of this kind are the Michelson or Fabry and Perot interferometers, the Echelon diffraction grating, or the Lummer-Gehrcke plate (15). The method of measuring the widths of lines with the interferometer consists in gradually increasing the distance between the plates until the interference fringes vanish, the "half-widths" of the lines being deducible from the optical difference of path when this occurs (13). By somewhat different methods the Echelon and Lummer-Gehrcke plate may also be used for measurements of the widths of lines. These instruments are of great value in all cases in which phenomena depending on small changes or differences of wave-length are concerned, though the largest gratings can also be used.

§ (7) The Zeeman Effect.—Amongst the phenomena included in this category may be mentioned the magnetic resolution of spectrum lines into components discovered by Zeeman in 1896 (16). Zeeman found that when a source of light is placed in a powerful magnetic field the lines are split up into components, some of which are polarised in a direction perpendicular to the magnetic field, whilst others are polarised parallel to the field, when the luminous source is observed transversely to the field. In the simplest case a line observed transversely is split up into three components, the outer components being polarised at right angles and the central component parallel to the lines of force. The separation of these two outer components is an important constant, from which the ratio of the electric charge to the mass of the electron has been deduced. It is found that $Zn = (\Delta\lambda/\lambda^2 H) \times 10^5$, where $\Delta\lambda$ is the separation of the outer components of a "normal" Zeeman triplet, λ the wave-length, H the magnetic field, and Zn a constant which, according to the most recent results, is equal to 9·385. When viewed along the lines of force the resolution in the simplest case is into two components of equal intensity which are circularly polarised in opposite directions. This simple case has been accounted for theoretically by Lorentz, but the phenomena are generally much more complex, many different types of resolution being now recognised. It has, however, been found that all lines belonging to the same spectral series exhibit the same resolution, and corresponding lines of different elements behave in the same way. Runge has shown that in many cases these complex separations may be expressed as simple fractions of the separation of the normal Zeeman triplet, but there are many complex types which do not appear to conform with this rule. The phenomena are even more complicated in the case of certain lines which exhibit dissymmetries and shifts, and complex phenomena are frequently observed in the resolution of lines which are themselves close doubles.

Many spectrum lines, when examined under a high resolving power, are found to possess a complex structure. This is most common in lines of the heavier elements, which are frequently accompanied by satellites. Certain regularities in the positions of these satellites have been observed, but their precise nature is not yet understood.

§ (8) Variations of Wave-length.—It has been found that the wave-lengths of lines are in many cases affected by the pressure of gas in the luminous source. The change in wave-length observed is usually very small, even when very great ranges of pressure are employed. These displacements appear to be directly proportional to the pressure and vary for different lines and for different elements. They are independent of the partial pressure of the luminous gas, but depend only on the total pressure. Recent investigations have shown that the pressure effect is an exceedingly complex problem.

It has been found that in electric arcs at atmospheric pressure minute changes of wave-length of some of the lines are to be found in the neighbourhood of the poles. This phenomenon, which is known as the Pole effect, is distinct from the change in wave-length due to pressure, and is of fundamental importance in the choice of standards of wave-length. The exact nature of the phenomenon is at present not fully understood.

Spectroscopic apparatus of very high resolving power is required in the investigation of the minute differences in wave-length which have been observed between the lines of the isotopes of lead.

§ (9) The Stark Effect.—Stark discovered in 1913 (17) that when certain spectrum lines are emitted in strong electric fields they are resolved into components, which, as in the case of the Zeeman effect, are polarised in different planes. This phenomenon differs from the Zeeman effect in the fact that the resolution is not the same for lines of the same series, the separation and number of the components increasing with the term number, though, as in the case of the Zeeman effect, the polarisation and number

of the components depend on the direction with respect to the field in which the observations are made. Stark's method of investigating the phenomenon was to observe the radiation of positive rays in a space between two metal plates which were kept at a suitable difference of potential in a specially designed vacuum tube, whilst Lo Surdo, who discovered the phenomenon independently at about the same time as Stark, found that the electric field in a vacuum tube in which the cathode was in a constricted part of the tube was great enough to exhibit the phenomenon. It has recently been found possible to investigate the electric resolution of the less volatile elements.

In some cases the resolution is symmetrical, whilst in others there is a marked dissymmetry of the components, both as regards their position and intensity. Lines of the diffuse series are affected to a much greater extent than those of the principal and sharp series. Stark has pointed out that the broadening of spectrum lines which occurs when powerful electric discharges are used is greatest for those lines which exhibit the greatest electrical resolution, and has suggested that the broadening in this case is due to the electric fields of neighbouring charged atoms on the radiating particles. Later investigations have confirmed this explanation of the broadening both qualitatively and quantitatively. It should be pointed out that this broadening by powerful condensed discharges, and also the broadening which results from an increase of pressure, is distinct from and superposed on the broadening which is due to the motions of the radiating particles in the line of sight. Whilst the latter effect imposes an inferior limit to the widths of spectrum lines, the former conditions can be varied from a negligible effect to a state in which the lines broaden to an almost continuous spectrum.

§ (10) Spectrum Series. (i.) *Balmer.*—It had been recognised for many years (18) that the lines in spectra were arranged in a manner which implied some definite relation between them, but the first successful representation of a series of lines was the formula of Balmer, which represents to a high degree of accuracy the chief series of lines in the spectrum of hydrogen. The word series is taken to mean a succession of lines the intervals between which become less from the red towards the violet and which degrade in intensity in the same direction. These series may consist of single lines, pairs or groups of more complex structure, and a spectrum usually comprises a number of such series. In the simplest case, that of hydrogen, Balmer expressed the wave-lengths of the lines (in reality complex lines) by the formula

$$\lambda = 3646 \cdot 14 \frac{m^2}{m^2 - 4},$$

where m takes a series of integral values 3, 4, 5, etc. The series therefore converges to a limit at $\lambda = 3646 \cdot 14$, but the limit of a series is never observed owing to the rapid degradation in the intensities of the lines, though in the case of hydrogen upwards of 30 members of the series have been observed in the chromosphere of the sun. In all the more recent work on spectrum series it has been found convenient to express the results in terms of the wave number, which is the reciprocal of the wave-length or the number of waves per centimetre. In terms of the wave number Balmer's formula reduces to

$$n = 27418 \cdot 75 - \frac{109675}{m^2}.$$

(ii.) *Rydberg, Kayser and Runge.*—Further regularities in spectra were recognised in the classical work of Rydberg and of Kayser and Runge. These investigators recognised the existence of three chief types of series which are now known (following Rydberg) as the Principal, Sharp, and Diffuse series, which are abbreviated P, S, and D respectively. Kayser and Runge expressed the wave numbers of the lines of series by the formula

$$n = A - Bm^{-2} - Cm^{-4},$$

where A, B, and C are constants, A being the limit of the series; but this method of representing the series is inferior to that of Rydberg, which not only expressed with success the wave numbers of the individual lines of series, but brought out the relation between the different series. Rydberg employed the formula

$$n = n_0 - \frac{N}{(m + \mu)^2},$$

where n_0 is the limit of the series, μ a constant peculiar to the series, and N the "universal constant," which is equal on Rowland's scale of wave-lengths to 109675, m taking successive integral values. A recent investigation of Curtis, in which the lines of the Balmer series of hydrogen were remeasured with great precision, has shown that on the international scale of wave-lengths the value N 109678·3 should be adopted. This "constant" is nearly the same for all series and all elements, and its interpretation in the light of the quantum theory is one of the most striking achievements of modern theoretical physics. The P, S, and D series differ physically in many important characteristics. Thus, of the three series, the P series, which usually contains the most prominent lines in the spectrum, shows the phenomenon of reversal most readily; in the S series, as its name implies, the lines are sharp under most conditions and are seldom reversed; whilst the lines of the D series are usually diffuse, and undergo a greater resolution in the electric field than the lines of the other series. Rydberg and Schuster discovered in-

dependently the important relation that the difference between the limit of the P series and that of the S series is equal to the wave-number of the first P line. The relations discovered by Rydberg can be seen from the following formulae, which represent the wave-numbers for the three series in the case in which each component of the series is a pair. Writing p, s, and d for the values of μ in the P, S, and D series respectively, and noting that the limit of the series n_0 in the simple Rydberg formula can be written in the form $N/(m+\mu)^2$, we have

P series,

$$n_1 = N\left(\frac{1}{(1+s)^2} - \frac{1}{(m+p_1)^2}\right),$$

$$n_2 = N\left(\frac{1}{(1+s)^2} - \frac{1}{(m+p_2)^2}\right);$$

S series,

$$n_1 = N\left(\frac{1}{(1+p_2)^2} - \frac{1}{(m+s)^2}\right),$$

$$n_2 = N\left(\frac{1}{(1+p_1)^2} - \frac{1}{(m+s)^2}\right);$$

D series,

$$n_1 = N\left(\frac{1}{(1+p_2)^2} - \frac{1}{(m+d)^2}\right),$$

$$n_2 = N\left(\frac{1}{(1+p_1)^2} - \frac{1}{(m+d)^2}\right),$$

n_1 in each case denoting the member of the pair of greater wave-number. The following important facts emerge from these relations:

(1) The S and D series converge to the same limit, but the separation of the components of the pairs remains constant.

(2) In the P series the separation of the components of the pairs decreases with the term number, converging to a common limit.

(3) If the more refrangible member of the pair is the stronger in the P series, the less refrangible member will be the stronger in the S and D series.

(iii.) *Ritz and Hicks.*—These relations are probably exact, though the wave-numbers of the individual lines of series may be more accurately represented by the formulae of Ritz and Hicks. The formula of Ritz may be written approximately

$$n = A - \frac{N}{(m+\mu+(a/m^2))^2},$$

while Hicks finds that the most accurate representation of the wave-number is given by

$$n = A - \frac{N}{(m+\mu+(a/m))^2}.$$

In the Ritz formula m takes integral values for the P and D series, but in the S series it assumes the values 2·5, 3·5, 4·5, etc. In the Ritz and Hicks formulae the values of μ and a for the P, S, and D series are expressed by the symbols $p\pi$, $s\sigma$, $d\delta$ respectively, and the relations between the series can be conveniently expressed by adopting the abbreviated forms due to Paschen, who writes, for example, mp, ms, and md, for the variable parts of the P, S, and D series; similarly $1p$ denotes the common limit of the S and D series. Similar relations to those given above are found in the case of series consisting of triplets. The above formulae require some amplification in the case of series which have satellites associated with the D series. It will be seen that every line in the scheme given above is represented as the difference between two wave-numbers, one of which is the limit of the series, and Ritz's important combination principle states that lines are often observed which can be represented by combining these wave-numbers in different ways. Thus in certain "combination" series, as they are called, the wave-numbers are equal to the differences between other observed wave-numbers. The validity of the combination principle has been established by the recognition of a great number of combination series. Of these combination series the most important is known as the Fundamental or F series, the wave-numbers of which can be expressed, in Paschen's notation, as $n_F = 2d - mf$. Ritz has suggested another representation of this series, but this does not appear to be fully established. The spectra of elements in the same group in the periodic table are similar, the separation of doublets or triplets being roughly proportional to the squares of the atomic weights.

§ (11) Arc and Spark Spectra (Fowler).—In a previous paragraph the terms arc and spark spectra were used to denote the lines which were relatively stronger in the spectra of electric arcs and spark discharges respectively. These terms have now assumed a new and more fundamental significance in the light of the important work of Fowler (19) on series in spark spectra. The fundamental difference in the spark and arc series lies in the fact that the "universal constant" N in the series formula is replaced by 4N in spark series, which is in accordance with recent theoretical investigations on the origin of spectra. The case of magnesium may be cited as an example. In the spectrum of this element there is a complete set of arc series with the "universal constant" equal to N, comprising a triplet system and a single line system, and also a pair system of spark lines in which the N is replaced by 4N. Both these sets of series are accompanied by combination series. These results are of fundamental importance in recent work on the origin of spectra in relation to atomic structure, but their discussion is beyond the scope of the present article.

§ (12) Band Spectra.—The existence of regularities is very apparent in the case of band spectra. Bands, which are usually associated with the spectra of compounds or molecules, consist of groups of lines which converge to definite heads, the head of a band being frequently, but by no means invariably, the strongest line in the band. These bands may be degraded either towards the red or towards the violet, and a large number of the lines composing them can often be enumerated, and bands of similar structure recur at intervals in the spectrum. Deslandres has shown that the wave-numbers of the lines composing bands may be expressed by a formula of the type $n = A + Bm^2$, where A is the wave-number of the head of the band, B a constant, and m takes a series of integral values 1, 2, 3, etc. In the same way it has been shown that the heads of the different bands may be arranged in series by a similar formula. The two formulae may be united into one, so that the complete representation of the lines in a series of bands is given by

$$n = Al^2 + Bm^2 + C,$$

where A, B, and C are constants and l and m are integers. These formulae are only approximate, and a number of more complex modifications have been proposed; but in the absence of any theoretical knowledge as to the origin of band spectra, it may be considered doubtful whether they are to be regarded as having greater significance than pure interpolation formulae. Band spectra are often exceedingly complex, several series frequently converging to the same head, and these different component series may show variations in their relative intensities under different physical conditions of excitation. Fowler (20) has recently observed a new type of band series associated with the spectrum of helium. In these bands the individual lines composing the bands are arranged in a manner which can be expressed approximately by Deslandres' formula, whilst the heads of the bands are arranged in series resembling those found in line spectra.

T. R. M.

References

Reference may be made to the following works on Spectroscopy, to which the numbered references given below are supplementary:

Handbuch der Spektroscopie, H. Kayser (Hirzel, Leipzig); *Spectroscopy*, Baly (Longmans, Green & Co., 1912).

1. Meggers, *Astrophys. Journ.*, 1918, xlvii. 1.
2. Lyman, *The Spectroscopy of the Extreme Ultra-violet*, 1914 (Longmans, Green & Co.).
3. *Phil. Trans.*, 1880, clxxi. 563.
4. Gouy, *Ann. Chim. et Phys.*, 1879, xviii. 5.
5. Cf. Pollok, *Practical Spectrographic Analysis*.
6. *Astrophys. Journ.*, 1908, xxvii. 296.
7. Schenck, *Astrophys. Journ.*, 1901, xiv. 116.
8. Schuster and Hemsalech, *Phil. Trans.*, A, 1900, cxciii. 189.
9. King, *Astrophys. Journ.*, 1908, xxvii. 353, and numerous later papers in *Astrophys. Journ.*
10. R. W. Wood, *Physical Optics*, 1914 (Macmillan).
11. *Phil. Trans.*, A, 1917, ccxvii. 237; *Proc. R.S.*, A, 1921, xcix. 78.
12. J. J. Thomson, *Rays of Positive Electricity*, 1913 (Macmillan).
13. *Phil. Mag.*, 1915, clxx. 274.
14. Buisson and Fabry, *Journ. de Physique*, 1912, ii. 442.
15. Gehrcke, *Anwendung der Interferenzen*, 1906.
16. Zeeman, *Magneto-optics*, 1913 (Macmillan).
17. *Elektrische Spectralanalyse*, 1914 (Hirzel, Leipzig).
18. Fowler, *Report on Spectrum Series*, Physical Society Publication.
19. Bakerian Lecture, *Phil. Trans.*, 1914, ccxiv. 225.
20. *Proc. Roy. Soc.*, A, 1915, xci. 208.

Spectrum: the arrangement of radiant energy in order of wave-length or frequency. See "Spectrophotometry," § (2).
Arc, the production of an. See "Spectroscopy, Modern," § (3) (ii.).
Flame, the production of a. See *ibid.* § (3) (i.).
Spark, the production of a. See *ibid.* § (3) (iii.).
Tube, the Production of a. See *ibid.* § (3) (iv.).

Spectrum Theory, application of quantum theory to. See "Quantum Theory," § (7).

Spherical Aberration: a defect of an optical system in which the focal length varies for different zones of the aperture. See "Microscope, Optics of the," § (5); "Telescope," § (3); "Optical Calculations"; "Lens Systems, Aberrations of."
In telescopes. See "Telescope," § (3).
Investigated by Foücault's original knife-edge method. See "Optical Parts, The Working of," § (3) (ii.).
Measurement of. See "Camera Lenses, Testing of," § (2); "Telescope," § (5).
Of eye. See "Eye," § (24).
Physical aspect of. See "Microscope, Optics of the," § (7).

Spherometer: an instrument for measuring the curvature of the surfaces of lenses and mirrors. See "Spherometry," §§ (1) to (6).

SPHEROMETRY

§ (1) In the following paragraphs methods will be described which are suitable for the determination of the curvatures of the surfaces of lenses and mirrors. Only those arrangements will be treated which are of general applicability, and which will be found useful in determining the structural data of optical instruments. A complete compendium of all the arrangements employed by different experimenters would be quite beyond the space at our disposal; but the methods described will be found to cover satisfactorily the whole range of curvature encountered in practice,

from the minute radii of the high-power microscope objective to the long radii of nearly flat surfaces.

I. Mechanical Methods

§ (2) The Simple Spherometer.—For medium curvatures, provided the diameter of the surface is reasonably large, the determination of curvature is most easily performed with a spherometer. The simplest form of spherometer is shown in *Fig.* 1. Three rigid legs, terminating in points a, b, and c, are attached to a stout plate. Through the centre of the plate passes a micrometer screw, also terminating in a point d.

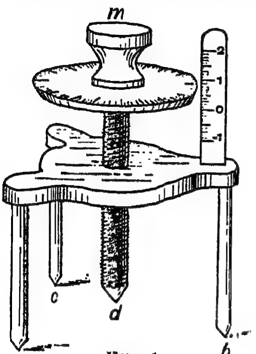


Fig. 1.

The three points a, b, and c are at the corners of an equilateral triangle, and d is equidistant from each of them. If the spherometer rests on a plane surface and d is screwed down so as just to touch the surface, the reading of the micrometer corresponding to zero curvature is obtained. If the instrument rests on a curved surface, and d is adjusted to make contact, the difference between the micrometer reading and that corresponding to zero curvature is a measure of the curvature. Suppose this difference is h, and that the radius of the circle on which the three fixed feet lie is ρ. In *Fig.* 2, $de = h$ and $eb (= ec = ea) = \rho$. From the geometry of the figure $ed \times ef = eb^2$, i.e. $h(2r-h) = \rho^2$, where r is the radius of the spherical surface. Therefore

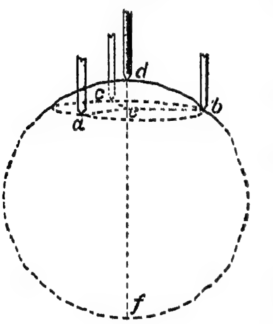


Fig. 2.

$$r = \frac{\rho^2}{2h} + \frac{h}{2}.$$

In order to determine the exact moment at which d makes contact, the usual method of using the instrument is to grip the milled head m, *Fig.* 1, lightly between the thumb and the first and second fingers, and screw up the micrometer, taking care to avoid bearing more heavily on one of the feet than on the others. So long as d is not touching, the spherometer remains quite steady; but as soon as d touches the surface any further rotation tends to raise the other feet from the surface and the whole spherometer rotates. With a little practice in handling the instrument, readings can be repeated to a surprising degree of accuracy by this method provided the span, 2ρ, is fairly large. This type of instrument is very useful therefore for large surfaces; but for surfaces of smaller diameter, where an instrument with the feet closer together must be employed, the precision of setting falls off considerably.

For the great majority of lenses the diameter is under two inches, and it is desirable to employ more accurate types of spherometer.

§ (3) The Abbé Spherometer.[1]—This spherometer, designed by Professor Abbé, gets over the difficulty of determining when contact is made by dispensing altogether with the micrometer screw and replacing it by a plunger which slides freely in guides and is pulled against the surface of the lens by a definite force. The degree of contact is therefore always the same. The position of the plunger is read by means of a microscope with micrometer eyepiece. *Fig.* 3 is a diagrammatic section of the instrument, from which its construction will be easily understood. B is a circular plate supported on two stout pillars, of which only one, P, is shown. The sliding rod R passes through the centre of B and through a lower guide G, which is attached between the two uprights. A counterweight W is attached to R by a string which passes over a pulley. This is slightly heavier than R, so that it tends to pull the latter upwards. A finely-divided scale is attached to the side of the plunger and is read by a microscope attached to the underside of P, fractions of a division being obtained by the micrometer eyepiece. The lens under test rests on a ring A which is concentric with the plunger. The ring is not bevelled to a single razor edge, but is ground to an annular plateau about $\frac{1}{2}$ mm. wide. Concave surfaces make contact with

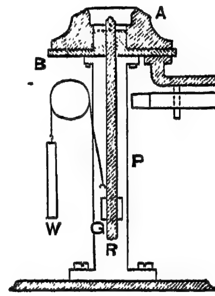


Fig. 3.

[1] See Pulfrich, *Zeits. f. Instrumentk.*, 1892, xii. 307.

the outer rim and convex surfaces with the inner. There is therefore a slightly different value for ρ, the radius of the ring, in the two cases. The reason for this arrangement is that with pointed feet, as in the simple spherometer first described, or with rings bevelled to a single razor edge, the points or edges become rounded in use and lenses of different curvature make contact at different distances from the centre. It is therefore impossible to ascribe an accurate value to ρ. With Abbé's arrangement the edges are much less acute and so do not round off readily. There are a number of rings such as A, but of different diameters, which may be fitted on the instrument so that lenses of various sizes may be measured.

§ (4) The Aldis Spherometer.—In this spherometer the lens rests on three steel balls instead of on points or rings. This appears to be the best method of support, inasmuch as the blunting difficulty is entirely eliminated; and, further, there is no tendency to scratch or mark the lens, a fault which some users allege against the ring support. With spherical supports the formula requires a little modification. Let AB, *Fig.* 4, be a convex surface of radius r resting on spheres of radius a whose *centres* are on a circle of radius ρ. The quantity which the spherometer measures is h, the vertical distance between the lowest point of the surface and the plane FG tangential to the spheres, the zero being obtained by the use of a flat surface in the ordinary way. This is equal to the vertical distance between the lowest point of A′B′, a curve concentric with AB but passing through the centres of the spheres, and the plane F′G′. Thus the ordinary spherometer formula, using the values h and ρ, would give the radius of A′B′ which equals the radius of AB plus the radius of the supporting spheres. Therefore

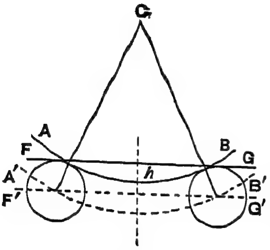


Fig. 4.

$$r=\frac{\rho^2}{2h}+\frac{h}{2}-a.$$

Similarly the formula in the case of a concave surface is

$$r=\frac{\rho^2}{2h}+\frac{h}{2}+a.$$

In the Aldis instrument, which is shown diagrammatically in *Fig.* 5, the lens is held down on the three-ball support by the weight of a plunger P, which slides smoothly in a hole in the superstructure. The micrometer screw is brought up from below. Before it makes contact the lens is held fairly firmly against the three peripheral spheres, and there is appreciable frictional resistance to rotation. When contact is made, however, the lens is practically held between the point of the micrometer and the plunger, and can be rotated freely.

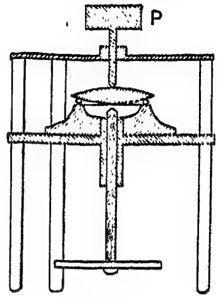


Fig. 5.

There are obvious defects in this arrangement for accurate work. The weight not only of the lens but of the plunger has to be supported on the sharp point of the micrometer. Quite an appreciable local depression is therefore produced before lifting occurs. This will occur to approximately the same extent in taking the zero reading, but the process is disastrous to any surface which undergoes it very often. The flat used for zero testing is soon ruined. The sensitivity with which the contact can be detected by rotating the lens is fairly high with large lenses, but for small lenses, such for instance as those in eyepieces, it becomes very low.

§ (5) The Guild Spherometer.[1]—This instrument is similar to the Aldis instrument as regards the arrangement of the micrometer parts and the use of three-sphere supports, but the method of detecting contact is quite different.

The micrometer screw, instead of terminating in a point, terminates in a small sphere of quartz (or glass). When this is nearly in contact with the surface of the lens under test, Newton's rings can be seen if a suitable arrangement for observing them be provided. From the behaviour of these rings as the micrometer is rotated the point of contact can be observed. The instrument is shown diagrammatically in *Fig.* 6. L is the lens under test resting on the three-sphere support (only two are shown). B is the small transparent sphere in which the micrometer terminates. A microscope M, with an ordinary vertical illuminator V fitted above the objective, is arranged for observing the Newton's rings formed between B and L. Illumination is supplied by a small 4-volt lamp on a suitable bracket. A piece of Wratten No. 25 gelatin filter (red) is mounted behind the aperture of the illuminator, and renders the

[1] Guild, *Trans. Opt. Society*, 1918, xix. 103.

light sufficiently monochromatic for several rings to be seen surrounding the point of contact.

The method of using the instrument is simple. Suppose that the small sphere B is gradually approaching the surface of L. When it comes within a short distance of it the Newton's rings become visible on suitably

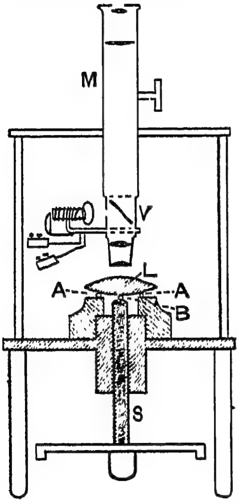


FIG. 6.

focussing the microscope. As the surfaces approach still closer the fringes expand outwards, a fresh one appearing at the centre for each half wave-length that the screw advances; but when actual contact is made the expansion of the fringes ceases abruptly, any further motion of the screw simply lifting the lens from its support. This point can be determined with certainty to less than one five-thousandth of a millimetre. It is better, however, to use as the criterion of contact not the point at which the expansion of the ring system ceases, but some easily recognised configuration *just before* this stage is reached. A convenient configuration is that in which the central black spot approximately trisects the diameter of the first black ring. This has the advantage that the appearance of the rings continues to alter for a little beyond the required setting, which is more satisfactory in practice than when the criterion adopted is one-sided. But the greatest advantage is that the surfaces are not in tight contact at this stage, and so no errors due to local distortion of the surface can arise.

With this instrument, exactly the same

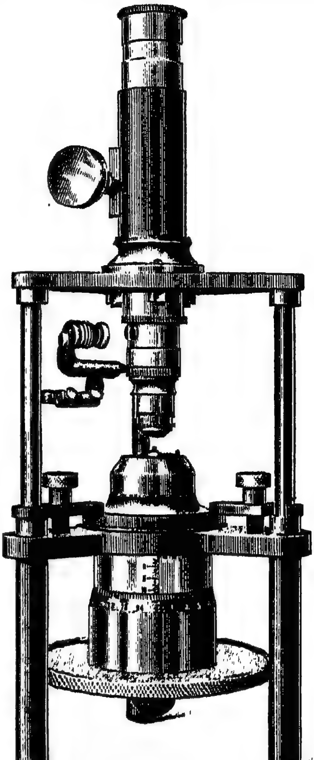


FIG. 6A.

accuracy in the micrometer settings is obtained with small lenses as with large. With a little practice readings of h can be repeated to a ten-thousandth of a millimetre, which is several times as sensitive as the best of the other existing types. An actual instrument is shown in *Fig.* 6A.

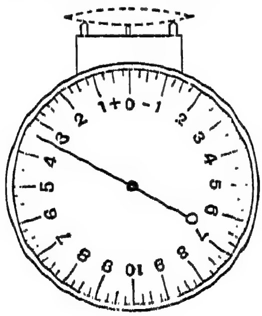


FIG. 7.

§ (6) THE WATCH-POCKET SPHEROMETER.—For very rough work, and for the rapid identification of surfaces, it is convenient to use a simple arrangement sold under the name of Dioptre-meter. In appearance it is similar to a watch, with three feet projecting from the rim (*Fig.* 7). The outer feet are fixed, and the central one can be pushed in against the pressure of a spring. When the feet are pressed against a lens surface, the central leg

is pushed inwards to an extent depending on the curvature of the surface. Its motion is converted into rotation of the dial pointer by toothed wheels.

As the instrument is principally used to measure the power of spectacle lenses, it is graduated to read the power of the surface in dioptres, assuming a refractive index of 1·52. Thus the actual reading gives the quantity $0{\cdot}52/r$, where r is the radius.

II. Optical Methods

§ (7) Newton's Rings.—For lenses at least 1 cm. in diameter, provided the curves are not too shallow, the curvatures can usually be most conveniently determined by means of an accurate spherometer. There are cases, however, in which the spherometer is inapplicable. For very small lenses, such as may be encountered in a microscope objective, or for lenses of shallow curvature for which the value of h would be very small, other methods have to be adopted. There are a great variety of such methods, but they are not all of satisfactory accuracy or convenience to be of practical value. We shall restrict ourselves here to a description of some of the methods which are of real utility in the optical laboratory. Of these the Newton's Rings method is the most suitable of any for convex curves of long radius; but by employing suitable arrangements it can be used down to quite small radii.

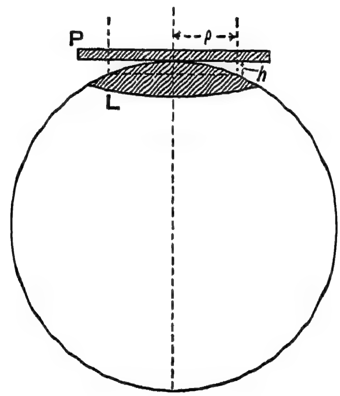


Fig. 8.

In *Fig.* 8 let P be a plate, of which the under side is accurately flat, resting on the surface of a lens L, and let a beam of monochromatic light fall on it from above. The lower surface of the plate forms with the lens an air film of which the thickness increases in all directions from the point of contact. Circular interference fringes surrounding the point of contact can therefore be seen if suitable observing apparatus is provided. Let us assume that the thickness of the film is zero at the centre. Since the ray reflected from the lower surface of the flat is incident from the glass side of a glass-air surface, it undergoes no phase change on reflection. The light reflected from the surface of the lens, being incident from the air side of a glass-air surface, undergoes a phase change of half a wave-length. Consequently, at the point of contact, where the thickness of the air film is zero, the rays annul each other by interference and produce darkness. There is therefore a black spot at the centre of the system. At any point away from the centre the illumination will be a maximum or zero according as the total retardation, which is equal to twice the distance between the surfaces + half a wave-length, is an even or odd number of half wave-lengths; that is, there will be a bright or a black ring at a distance ρ from the centre according as $2h$, *Fig.* 8, is an odd or even number of half wave-lengths. For purposes of measurement the black fringes are always used. These will be at such distances from the centre that twice the separation of the surfaces $=m\lambda$, where λ is the wave-length and m has successively the values 1, 2, 3, etc. If ρ_n is the radius of the nth ring, $2h$ for this ring $=n\lambda$. But $2rh-h^2=\rho_n{}^2$, where r is the radius of the curved surface. Since h is only a few wave-lengths, h^2 may be neglected in comparison with $2rh$, so that

$$\rho_n{}^2=2rh=rn\lambda,$$

$$\therefore \rho_n=\sqrt{rn\lambda}.$$

The radii of the rings are therefore proportional to the square roots of the natural numbers, and the fringes become closer together as we pass outwards from the centre. The width of a fringe is very approximately inversely proportional to its distance from the centre; for if

$$\rho_n{}^2=nr\lambda,$$

$$\rho_{n-1}{}^2=(n-1)r\lambda,$$

$$\therefore (\rho_n-\rho_{n-1})(\rho_n+\rho_{n-1})=r\lambda.$$

Approximately, therefore, the width of a fringe

$$=\rho_n-\rho_{n-1}=\frac{r\lambda}{2\rho_n}.$$

It is easier to measure the diameter of the rings than their radii. If D_n is the diameter of the nth ring,

$$r=\frac{D_n{}^2}{4n\lambda}. \quad . \quad . \quad . \quad . \quad (1)$$

This relation can be employed to determine the radius of curvature of the surface. It assumes, however, that the flat is geometrically tangential to the lens at the point of contact. In practice the nature of the contact is by no means definite. If the surfaces were to touch at one point only the pressure would be

infinite. There is always, therefore, a flattening of the curve and denting of the flat to spread the contact over an appreciable area. Owing to this deformation near the middle of the system the diameters of the rings will not be accurately given by the formula we have discussed. However, the *difference* in the diameters of rings at some distance from the centre will not be affected, and the radius of the surface will be correctly given by

$$r = \frac{D_{n2}^2 - D_{n1}^2}{4(n_2 - n_1)\lambda}, \quad . \quad . \quad . \quad . \quad (2)$$

where D_{n2} and D_{n1} are the diameters of the n_2th and n_1th rings respectively. In carrying out this method for lenses of long radii a travelling microscope, such as that employed for the measurement of spectrum photographs, etc., is essential. The lens is laid on the stage of the microscope, *Fig.* 9, with the surface to be tested upwards, and the flat plate laid on top of it. Unsteadiness can be prevented by little pellets of plasticene or soft wax near the periphery. The illumination should be supplied in a parallel beam at exactly normal incidence by means of an arrangement of the type shown in the diagram. A is a small aperture in a screen, behind which is placed a strong source of monochromatic light. The aperture is at the focus of a lens L. The light is reflected at a transparent glass reflector R, which is mounted in a bracket with universal adjustment attached to the framework which carries the microscope. By adjusting R the light can be brought to normal incidence.

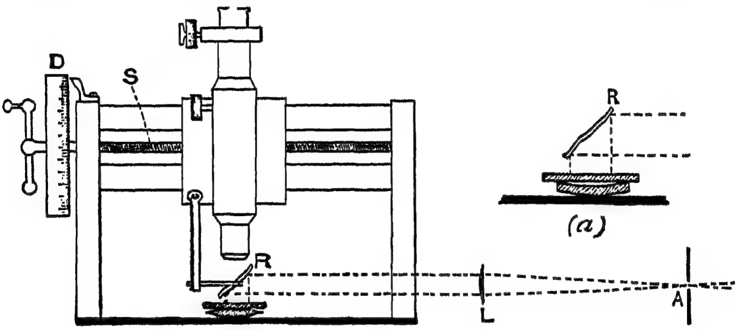


FIG. 9.

In making the actual measurements it is necessary, as we have seen, to measure the diameters of two rings, the n_1th and the n_2th. The accuracy of the final result will be greater the greater $n_2 - n_1$. It is not desirable, however, that *either* n_2 or n_1 should be small, as the rings near the centre being broad and diffuse the accuracy of setting is considerably less than for rings further out. The effect on the result of errors in D_{n2} and D_{n1} is readily deduced by differentiating equation (2) with respect to each of them separately. It requires D_{n2}/D_{n1} times as big an error in D_{n1} as it requires in D_{n2} to produce a given error in r. Thus the relative accuracy with which the diameters have to be measured is proportional to the diameters. We saw that the widths of the fringes were inversely proportional to their diameters, so that if settings were accurate to the same fraction of a fringe width the probable error in the result would be of the same order in each measurement. As a matter of fact the accuracy of setting does vary nearly in this way at some distance from the centre, but for the first few fringes the fringe width is varying rapidly and the judgement of the setting is affected by the want of symmetry. It is not desirable therefore to make measurements on rings of smaller order than the 6th or 7th. For the outer ring as large a value of n as possible should be taken. This involves working where the fringes are very close together, and it will be found of convenience to employ a micrometer eyepiece to the microscope. The microscope as a whole is moved slowly across the fringe system by means of the screw S, the fringes being counted as they pass the cross-lines. When the fringe is reached at which it is intended to stop, the reading on the drum D is taken. Several settings are then made by means of the micrometer eyepiece, and the mean of these readings used to correct the drum reading. The advantage of this process is that if the repeated settings are made with the screw S, the cross-lines have to be brought back several fringes each time in order to take up the backlash of the screw. It is very easy in doing this to get confused as to the fringe on which the measurement is being made. If the repetitions are made with the micrometer eyepiece, however, the screw being of finer pitch and the mechanism being more delicate altogether it is not usually necessary to retreat more than a fringe width to take up the backlash, and the chances of getting on the wrong fringe are minimised.

There are one or two details in the arrangement of *Fig.* 9 which differ from those usually described in text-books and papers, and which require some explanation. The lens is universally made to lie on the flat surface, and in consequence the fringes are seen *through* the lens, and are in general magnified. An object in contact with one surface of a lens is magnified, when observed through the lens, in the ratio f/f_b, where f is the focal length and f_b is the "back focal length," *i.e.* the distance from the principal focus to the surface of the lens. It is only in the special case of a lens in which the other surface is flat, and in which, therefore, the appropriate nodal point coincides with the curved surface, that $f_b=f$. In all other cases the distance traversed by the microscope has to be multiplied by f_b/f to give the true diameters of the rings. In the diagrams illustrating the method this point is always avoided by the convenient choice of a plano-convex lens; but in practice this is the exceptional case. The expedient of placing the flat uppermost, as in *Fig.* 9, overcomes this difficulty entirely.

The use of the illuminating arrangements described here, instead of the usual vertical illuminator above the microscope objective, ensures a parallel beam at normal incidence for all positions of the microscope. With the ordinary method of illumination, in which the light comes as a converging cone from the objective, fringes are only visible with extremely thin air films, and consequently the lens and flat must be in contact. This being so, the expedient is adopted, in dealing with concave surfaces, of measuring the difference in curvature between the surface and a convex surface of greater curvature which rests within it. This is an inaccurate arrangement except when the two curvatures are very nearly equal; as an error in the assumed value of the radius of the convex surface produces a greater percentage error in the result obtained for the concave.

Suppose h is the vertical distance between the two surfaces at a distance ρ from the centre. Let h_1 and h_2 be the respective distances of the convex and concave surfaces above a plane which is tangential to them at their point of contact.

Then if r_1 is the radius of the convex and r_2 that of the concave surface

$$\frac{1}{r_1}=\frac{2h_1}{\rho^2},\quad \frac{1}{r_2}=\frac{2h_2}{\rho^2},$$

$$\therefore \frac{1}{r_1}-\frac{1}{r_2}=\frac{2}{\rho^2}(h_1-h_2)=\frac{2h}{\rho^2}\doteq\text{constant};$$

$$\therefore \frac{1}{r_1{}^2}dr_1=\frac{1}{r_2{}^2}dr_2$$

or

$$\frac{dr_1}{r_1}\cdot\frac{r_2}{r_1}=\frac{dr_2}{r_2}.$$

Thus a percentage error dr_1/r_1 in the assumed value of r_1 gives rise to a percentage error dr_2/r_2, which is r_2/r_1 times as great. Since, except in special cases, r_1 has itself to be determined, and cannot be known to a greater percentage accuracy than is also desired for r_2, the method is not a good one.

With the method of illumination shown in *Fig.* 9 it is quite possible to get fringes with a large total path difference and to measure the radius of a concave surface directly against a flat. The flat is laid on the rim of the lens, as in *Fig.* 9 (*a*), and when the light is properly adjusted the fringes will be seen and can be measured in the same way as for a convex surface. The light has to be absolutely normal, and the angular aperture which the source subtends must be small. The reasons for these precautions need not be treated here; they will be understood from the treatment of interference by thick plates in another article. In making the adjustments the best method of procedure is to mount the lens L on a stand which has rack-and-pinion adjustments both horizontally and vertically. Remove L and adjust R by hand until the fringes appear as distinct as possible. Replace L, and adjust it horizontally and vertically until the rings are equally distinct in all parts of the field. A guide to the correct adjustment will be found in the size of the rings near the centre. Since for normal incidence the retardation is a maximum, the rings shrink to a minimum size when the incidence is normal. The movements of the rings with slight changes in the adjustment of the light give rise to the impression that the results obtained depend entirely on the adjustment. This is not so, however, as the *relative* positions of different parts of the ring system are affected by the adjustment of the light to exactly the same extent as if the surfaces were in contact as in the ordinary method.

The necessity of employing as small a hole as possible at A necessitates the use of a very bright source of light. The green line of mercury is the most useful to employ, and should be obtained from a quartz arc made specially for interference work.[1] The other lines in the spectrum are much less intense than the green line, and only detract slightly from the blackness of the fringes, the positions of which are determined by the green line. The best type of aperture to use at A is a small iris diaphragm, which can be adjusted to give the best compromise between brightness and distinctness.

In employing Newton's Rings for medium and steep curves, the travelling microscope ceases to be satisfactory on account of the smallness of the radii of the rings. It is better in this case to use an ordinary microscope with

[1] Guild, *Proc. Phys. Soc.*, 1920, xxxii. 341.

a micrometer scale in the focal plane of a Ramsden eyepiece, and to measure up the rings on this scale. An objective should be chosen of such power as to give an image of the ring system of suitable size. With steep curves, the objective requires to be of too high a power to allow the interposition of the vertical illuminator shown in *Fig.* 9, and an ordinary vertical illuminator above the objective must be employed. The arrangement cannot then be used for concave surfaces. However, for medium and steep curves the method is no more accurate and is much more tedious than several other methods which will now be described.

§ (8) Microscope Methods.—The following methods require a good microscope stand with mechanical stage and a universal sub-stage with centring adjustments. A vertical illuminator of the ordinary inclined cover glass type is mounted above the objective.

(i.) *For Concave Surfaces.*—The lens is mounted on a support attached to the sub-stage fitting, which can be raised or lowered by the rack and pinion P, *Fig.* 10 (*a*), and centred by the screws *ss*. A piece of thin cover glass on which some easily recognisable markings have been scratched is placed on the stage, and the microscope is focussed on these, vertical illumination being supplied by the lamp L. The sub-stage is then racked up or down until the image of the markings formed by reflection at the concave surface is also seen in sharp focus. By means of the screws *ss* the image is centred with respect to the object, which is then at the centre of curvature of the surface. The cover glass is removed and the microscope racked down until some lycopodium grains sprinkled on the surface of the lens are in focus. The distance the microscope has been moved is the radius of curvature required. The method is accurate to about 0·05 mm.

(ii.) *For Convex Surfaces.* For this case the arrangement is somewhat more complicated. The microscope is initially racked down so that the objective is about the level of the stage. The sub-stage is adjusted until some lycopodium grains on the surface of the lens are in focus. The surface is then at the object point[1] of the microscope. A piece of silvered glass with a number of scratches cut on it is placed at O, and its distance from the vertical illuminator is adjusted until an image of the scratches is seen simultaneously with the lycopodium. Under these circumstances light from the scratches, after reflection at the vertical illuminator and passage through the objective, converges towards the object point of the latter. The sub-stage is then racked upwards until an image of the scratches is again in focus. This happens when the rays converging from the objective meet the surface normally — that is, when the object point of the microscope coincides with the centre of curvature of the surface. The microscope is now racked up-

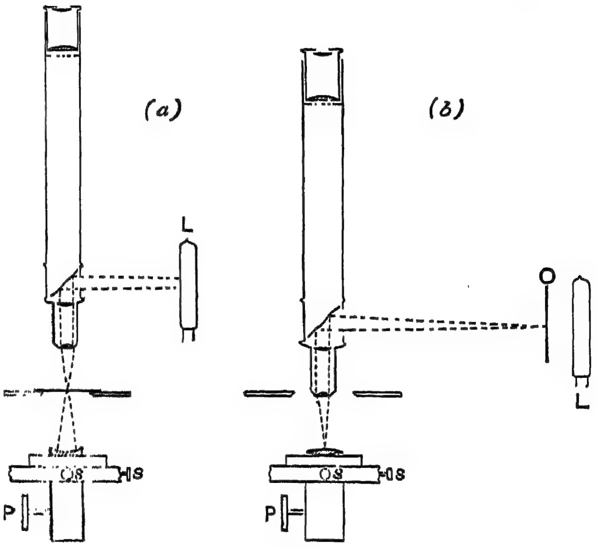


Fig. 10.

wards until the lycopodium is again in focus, and the distance moved is the radius of the surface.

This method is frequently described in an inaccurate manner. By following the procedure outlined above, it will be observed that all adjustments involving the image of the scratches are made with the microscope and the object O in a fixed relative position. This is clearly an essential condition to accuracy. The results are reliable to about 0·05 mm., as in the case of the previous method; but the range of curvature for which it can be used is limited by the working distance of the objective which is available. In the previous method the range was limited only by the range of vertical traverse of the microscope.

(iii.) *Magnification Method for very large Curvatures.*—For large curvatures such as are met with in some of the lenses of microscope

[1] For convenience of description we may use this term to denote the point conjugate to the centre of the focal plane of the eyepiece. To fix the position of this plane a graticule should be fitted.

objectives, none of the methods so far mentioned are satisfactory. The following method is at once extremely simple and extremely accurate; and has the great advantage that the percentage accuracy is practically independent of the curvature. Suppose (*Fig.* 11) that we employ as object a suitable mark on the front lens of the object glass of the microscope, and so adjust a reflecting surface M that an image of this object is seen in sharp focus in the focal plane of the eyepiece. It follows that the image I formed by reflection in the surface is situated at the object point of the microscope and is therefore at a fixed distance from O, $=k$ say, whatever the curvature of the surface may be. From the magnification of the image the curvature can be deduced. Let u be the distance from the surface to the object, v the distance from the surface to the image, and r the radius of the surface. Then $1/u+1/v=2/r$, quantities being regarded as positive when measured against the incident light. $k=\text{IO}=u-v$. The magnification of the image at $\text{I}=-v/u,=m$ say. Hence

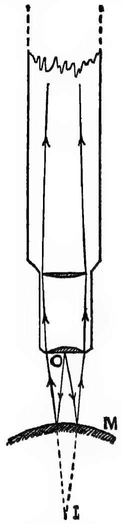


FIG. 11.

$$1-m=\frac{2v}{r},\ \ 1-\frac{1}{m}=\frac{2u}{r},$$

$$\therefore m-\frac{1}{m}=\frac{2k}{r} \quad . \quad . \quad . \quad \text{(i.)}$$

or

$$r=\frac{-2mk}{1-m^2}. \quad . \quad . \quad . \quad \text{(ii.)}$$

From (i.) we see that in general there are two distances of the mirror from the object which satisfy the condition of the experiment, namely the value of u which gives a magnification m, and another value u' for which the magnification is m', where $m'=-1/m$. In the case of a convex surface one of these values is negative and so is experimentally inadmissible. The other value is such that the magnification is $+ve$ and numerically less than unity.

For a concave surface, both values of u are positive. In one case u is less than $r/2$ and the magnification is positive and greater than unity. In the other case u is greater than r, and the magnification is $-ve$ and numerically less than unity.

This method can be employed over a large range of curvature; but the best experimental conditions differ for different ranges.

For very short radii the arrangement of *Fig.* 12 is convenient. The microscope is fitted with a vertical illuminator, preferably of the type shown, in which the reflecting cover-slip is just under the eyepiece. Light is supplied by a small 4-volt lamp, attached for convenience to the microscope. The illuminated aperture of the lens itself serves as the object. The curve C to be measured is placed centrally on the stage, and, on racking the microscope up or down, a circular disc will be seen which is the image by reflection in C of the aperture of the objective. This is sharply focussed and its diameter measured on a micrometer scale in the focal plane of the eyepiece. The actual dimensions depend of course on the magnification of the microscope as well as the magnification m of the reflected image. However, for a given objective and tube length, the measured diameter is proportional to the magnification m. The instrument is calibrated by measurements on a series of the steel spheres sold as ball-bearings. These can be obtained of diameters ranging from $\frac{1}{32}$ of an inch to 3 inches by small steps. The sizes up to about an inch make excellent optical standards, and a complete set should be available in any laboratory. They can be measured with a micrometer gauge; but it will be found that they are invariably so close to their nominal values that for optical purposes these values can be taken as substantially correct.

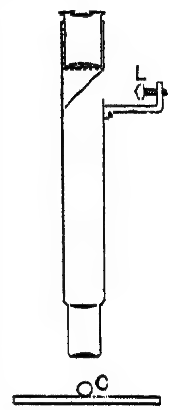


FIG. 12.

The accuracy with which the diameter of the luminous disc can be measured depends of course on its size. An objective should be employed which gives an image of at least 4 mm. on the eyepiece scale. The size of the image depends not only on the power of the objective but also on its working distance and the linear dimensions of the aperture. However, it is easy to find a series of three or four objectives which between them will cover the range from the smallest radii encountered in microscope work up to radii of about 10 mm. If, by means of the steel balls, calibration curves are drawn for each objective for the range in which it is to be used, they will be found to be practically straight lines within the accuracy of observation. This follows at once from the relation $r=-2mk/1-m^2$ when m is very small, as is the case with surfaces of such short radii. Thus the scale reading is proportional to the

radius. By a single reading the radius of any surface can be obtained from the calibration curve. It is safer, however, to make check measurements on one or two of the spheres each time the apparatus is reassembled, to ensure that the constant of the instrument has not been altered.

It is evident that the calibration curve obtained with the spheres will do equally well for the measurements of concave curves, the real image formed being of the same magnification as the image obtained with a convex curve of the same radius.

It will be observed that if suitable objectives are employed, the diameter of the disc to be measured is of the same order of magnitude (40 to 100 scale divisions), however small the radii tested may be. Since it is easy to read to ·1 scale division (with a good quality micrometer scale), the average sensitivity of the method is about 1 part in 700. Further, the method is virtually one of substitution, and all aberrational effects of the observing microscope are eliminated.

Thus there is no source of error other than the accidental errors of measurement, and the accuracy is equal to the sensitivity.

In practice the definition is improved by using a light green filter to cut out the extremities of the spectrum, as, if white light is used, chromatic aberration tends to spoil the sharpness of the image.

(iv.) *Magnification Method for Medium Curvatures.*—It is not convenient to use the method in the form described above for radii much over 1 cm., because the disc becomes too large to be contained in the field of view even with the lowest power objectives.

For larger radii it is better to use as object a scale, similar to that employed in the eyepiece, attached to the front of the objective with soft wax or plasticene. An image of this scale is seen when the object is at the correct distance from the surface. By measuring the dimensions of a suitable number of graduations, the magnification can be obtained from a knowledge of the magnification of the microscope. We may, if we choose, determine k, and calculate the radius directly from formula (ii.) above. To measure k a flat surface, preferably silvered, is laid on the stage and the microscope focussed on some scratches on the silver. On racking downwards, an image of the scale comes into focus. The distance which the microscope has been moved is clearly $\frac{1}{2}k$.

As in the case of very small radii, however, it is better to calibrate the instrument empirically. Unfortunately, there are no easily available standards such as the ball-bearings employed in the former case; but points on the curve can be obtained from measurements on lens surfaces which have been determined accurately with a spherometer. If sufficient of such standards are available to determine the calibration curve accurately, an unknown radius can be obtained directly from the curve; but it is usually better to measure m for the surface in question and also the corresponding value for the *nearest* standard, and deduce the unknown radius from the formula

$$r=r_1\frac{m(1-m_1^2)}{m_1(1-m^2)}.$$

In the case of concave curves, if the radius does not exceed two or three centimetres, it is easy to obtain the real and diminished image, and in this case the convex standards apply to concave curves as well. With shallower curves, however, the real image is too far from the lens to be within the range of vertical traverse of ordinary microscopes; and it is necessary to work with the virtual image, for which, as we saw earlier, m is greater than unity. A separate set of concave standards is necessary for those. Unfortunately, as the magnification increases the image of the scale becomes very faint, and the measurement is somewhat difficult with magnifications greater than unity. The range of curvature for which the diminished real image can be utilised may be increased by mounting the lens on the sub-stage or even on a separate support below the microscope.

Except with comparatively small radii the 4-volt lamp L (*Fig.* 12) will not give a sufficiently bright field. It is better to employ a 100 c.p. "Pointolite" lamp, or any similarly intense source, behind a ground-glass screen situated about 30 cm. from the microscope and on a level with the aperture of the illuminator. Except for the ground glass the lamp should be entirely enclosed.

The scale employed as object should not be too fine. A $1\frac{1}{2}$-inch objective with a working distance of about an inch will give satisfactory results over most of the range for which the method is applicable.

There are a variety of other magnification methods; but they involve the measurement of two magnifications and are inferior in convenience and accuracy to that just described.

§ (9) Direct Location of Centre.—For concave curves of radii above 20 cm. this method may be employed provided the ratio of the diameter of the surface to the radius of curvature is not too small. The principle of the method is that if an object and its image by reflection in the surface are coincident, or in close juxtaposition in a plane perpendicular to the axis of the surface,

they are located at the centre of curvature. The method as usually described, in which the image is formed on a card, is not, however, very sensitive; and if the best possible accuracy is required in the result the following arrangement will be found much better. A microscope slip is silvered or platinised on one side, and a network of very fine scratches is made with a sharp needle. The slip is mounted on the stage of a microscope which is turned back so that the microscope tube is horizontal. The scratches are illuminated by means of a small prism P, *Fig.* 13 (*a*), and a "pointolite" lamp S. The concave surface M is arranged so that an image is formed on the slip a little microscope more critically than almost any other type of object, and the depth of focus is reduced to a minimum. The accuracy depends on the ratio of diameter to radius of curvature. It is about 0·1 millimetre for a ratio of 1 : 20.

The direct location of the centre is unfortunately impossible with convex surfaces. The method in which an auxiliary lens is used to produce a convergent beam, which meets the convex surface normally and is reflected back through the lens and focussed in the vicinity of the object, is valueless for work of any accuracy on account of the aberrations of the lens.

§(10) Foucault's Shadow Method.—This

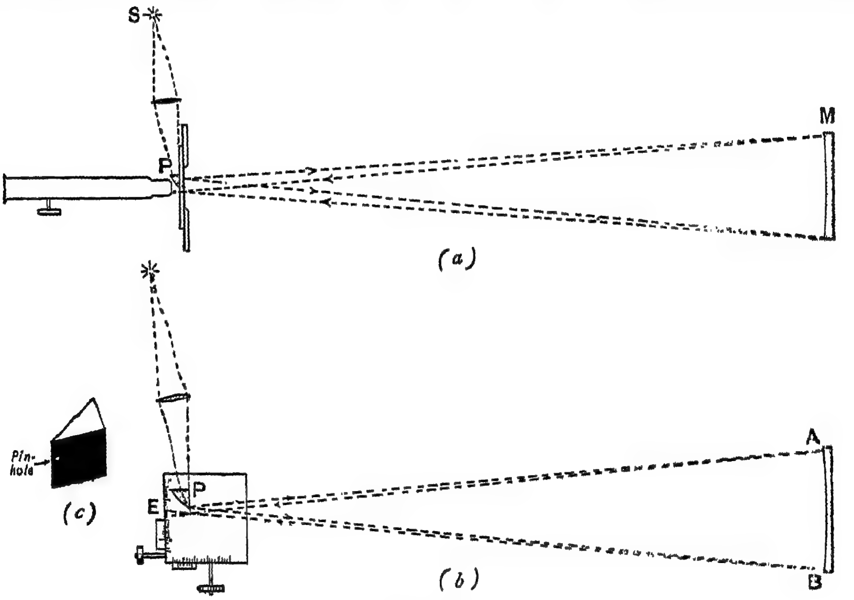


Fig. 13.

to the side of the illuminating prism P. In this region the plating should be almost all scraped off, a few streaks being left on which to focus the microscope. When the adjustment is correct, these streaks and the reflected image of the illuminated scratches will be simultaneously in focus. The plated side of the slip should be towards the mirror, and the distance between them can be measured in any convenient manner. It is advantageous in practice not to attempt precise adjustment of the mirror, but to get it approximately correct and measure the error by racking the microscope so that first the surface of the slip is in focus and then the reflected image. Half the distance between the two positions is the correction to be applied to the distance from slip to mirror. The edges of a silvered film can be focussed up with a is another useful method for concave curves. It forms at the same time a sensitive test of the sphericity of the surface. A convenient arrangement is shown in *Fig.* 13 (*b*). P is a small 45° prism with clean sharp edges. One of the square faces is covered with tinfoil in which a fine hole has been pierced with a needle close up to one of the 45° angles, *Fig.* 13 (*c*). This prism is mounted on a platform capable of fine adjustment in two directions. A good mechanical stage of a microscope will do. The pinhole is illuminated by focussing on it a very bright source, *e.g.* a "pointolite" lamp. A beam of light spreads out from the pinhole, and the mirror AB is placed so that an image of the pinhole is formed just a little to the side of the prism. The eye is placed just behind the prism at E so as to receive the reflected light. Since

this comes from all parts of the surface the latter appears uniformly bright all over. If, now, the prism is moved laterally, so as to bring the pinhole closer to the axis of the mirror, a point is reached at which the edge of the prism cuts into the reflected beam and shuts off some of the light reaching the eye. Assume first that the surface is truly spherical so that all rays corresponding to one point of the object converge to one point in the image. If P is exactly at the centre of curvature it will cut the beam exactly at the image, and the surface will darken uniformly all over. If P is inside the focus the rays from A are intercepted before those from B and a shadow comes over the surface from A to B. If P is outside the focus the shadow starts from B. The stage is moved longitudinally until the lateral movement produces a darkening of the whole surface simultaneously. The distance from the pinhole to the surface is then the radius of curvature. As a test of the surface this method is valuable. If there are any parts of it from which rays do not converge to the same point as from others, the surface will not darken uniformly in any position of P. If, for instance, the focus for the peripheral zone is farther out than for the central zone, and P is at the focus of the latter, the central zone will darken uniformly, but the rays from A will be cut off, while those from B will not. A crescent-shaped region near A will therefore remain bright after the remainder of the surface is obscured. If P is at the focus of the marginal rays the regions near A and B will darken simultaneously; but rays from the side of the central zone nearest A will pass the edge of the prism and this region will appear bright.

The effect of local irregularities is to give a patchy appearance to the field. The exact path of the rays from different parts of the surface can be determined, for the place where the rays from any two parts, say the centre and any other region, intersect the same vertical plane parallel to the axis is found by adjusting P till the two parts in question darken simultaneously. By making observations with AB in azimuths differing by 90° the paths of the rays may be completely determined.

§ (11) Summary. The methods of curvature measurement described in the preceding sections provide complete equipment for curvatures of all possible magnitudes. Where to leave off one method and use another depends largely on the predilections of the observer, as their ranges of utility overlap. As the intention was to give only those methods which are of general application, description has had to be omitted of various useful methods, suitable for works practice, where the possession of both the convex and concave curves of any radius is assumed.[1] J. G.

SPIRIT-LEVELS

§ (1) General.—Spirit-levels are used to determine the direction of the horizontal, or vertical, at any point, and sometimes to measure small angles from the horizontal. The usual form consists of a glass tube shaped so that the longitudinal section of its upper internal surface is the arc of a circle. The tube is nearly filled with liquid, so that a bubble of vapour is left. The bubble rises to the highest portion of the tube, and when at rest the line joining the two ends is horizontal. The tube is generally graduated on its upper surface. The graduations are generally single or double millimetres, $\frac{1}{10}$ths or $\frac{1}{20}$ths in., or in older instruments Paris lines (2·26 mm.), and may be continuous or have the centre marks omitted. The line joining the ends of the bubble, when those ends are equidistant from the centre of the tube as shown by the graduations, may be called the axis of the bubble. The glass tube is generally fixed in a metal tube, cut away to enable the bubble to be seen, and this tube is again adjustably fixed to some mount. For rough work it is sufficient to fix the bubble tube so that the instrument to which it is attached is level when the bubble is in the centre of its run, and trust to the adjustment remaining constant. The bubble is in adjustment when its axis is parallel to the base of its mount, or to the surface which it is desired to level. A tubular bubble will obviously only indicate the horizontal in the direction of its length, and if a surface has to be levelled two bubble tubes are required at right angles to each other.

§ (2) Circular Bubbles.—For rough work a circular bubble is often convenient. In this case the upper surface of the vessel containing the liquid is a portion of a sphere, and dislevelment in any direction is shown by the one bubble. *Fig.* 1 shows two forms of circular

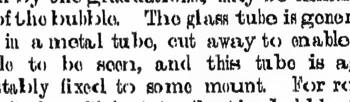


Fig. 1.

bubbles. A has the vessel containing the liquid formed of a completely sealed-up glass vessel, B has only the top of glass; it is more compact than A, but it is difficult to ensure that there shall be no slow leakage.

§ (3) Tubular Bubbles.—Where the horizontal is required with an error of less than a minute or so of arc, tubular bubbles are

[1] See, for instance, S. D. Chalmers, *Trans. Opt. Soc.*, 1915-16, xvi. 168.

almost always used, and they must be made so that they can be reversed end for end in order to check the adjustment.

For levelling a plane surface the under surface of the level mount is also a plane surface. If this now rest on a flat surface, which is tilted as necessary till the bubble lies in the centre of its run, the surface is level in the direction of the length of the bubble tube, provided the bubble is in adjustment. The adjustment is checked by reversing the bubble tube end for end. If the bubble still lies in the centre of its run the adjustment is correct. If it has moved it is brought back half-way by tilting the surface and the rest of the way by the adjusting screws on the bubble mount. The whole is then checked by reversal. It is not necessary for the bubble to be in adjustment; if the bubble remain in the same position in the tube after

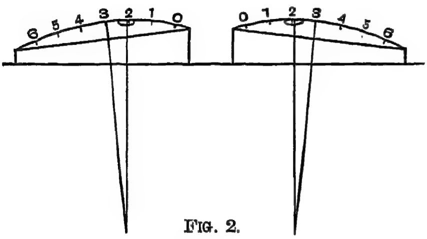


FIG. 2.

reversal, the surface is level. A consideration of *Fig.* 2 will make this clear.

§ (4) STRIDING LEVEL.—For levelling the axis of a cylinder, such as the telescope of a collimator or the horizontal axis of a transit, the level is mounted as shown at Q in *Fig.* 3. This is known as a striding level.

§ (5) USE OF LEVEL FOR MAKING A VERTICAL AXIS VERTICAL.—In order to make the vertical axis of rotation of any instrument (*e.g.* a theodolite) truly vertical, a bubble tube is fixed to the rotating portion of the instrument at right angles to the axis. If the bubble axis is truly at right angles to the vertical axis it is only necessary to set the bubble by means of the instrument levelling screws, in two positions at right angles to each other. In practice the instrument is rotated till the bubble tube is parallel to two of the levelling screws of the instrument, and by their means the bubble is set, *i.e.* brought to the centre of its run. The instrument is then rotated through 180°. If the bubble remain set the bubble is in adjustment; if not, the bubble is brought back half-way to its original position by means of the levelling screws, the instrument is again rotated through 180° into its original position, when the bubble should retain the mean position at which it was last set. The instrument is then rotated through 90° and the bubble set to the mean position as found above by means of the third levelling screw. The bubble should now take up the same position however the instrument is rotated, and if this is so the vertical axis is vertical. It is usually necessary to repeat the reversals several times before this condition is obtained. If the mean position of the bubble is not the centre of the tube the

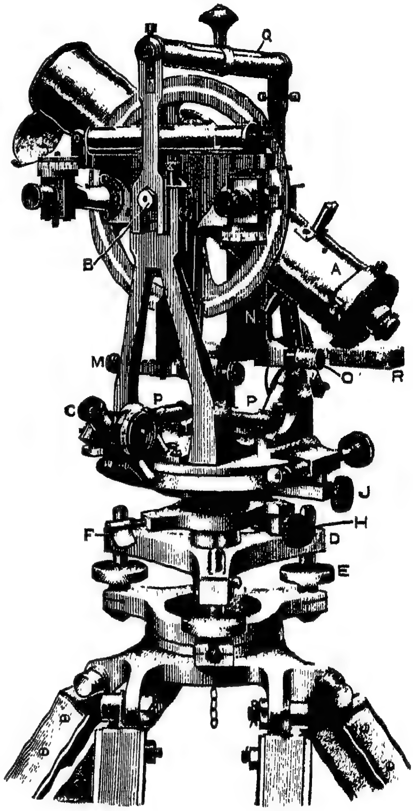


FIG. 3.

bubble can be brought into adjustment by means of the bubble-adjusting screws, but unless it is badly out of adjustment this is not worth while.

§ (6) THE SENSITIVENESS OF A BUBBLE, *i.e.* the amount the bubble will move when the tube is tilted through a small angle, varies directly as the radius of curvature of the tube, and is generally measured by the angle of tilt necessary to move the bubble through one division of the tube.

§ (7) THE ACCURACY OF A BUBBLE depends on other factors as well as on its sensitiveness, notably on the length of the bubble itself, and the method of illumination and viewing

adopted, and also of course on the accuracy of the figuring of the surface of the tube. When the tube is tilted and the ends of the bubble cease to be at the same level, a force due to this difference of level tends to move the liquid. This force is proportional to the specific gravity of the liquid and to the difference in level of the ends of the bubble. The difference of level or "head" is $l \sin a$, where l is the length of the bubble and a the angle of tilt. This force therefore varies as the length of the bubble, and is independent of the curvature. This force has to overcome the inertia of the liquid and the frictional forces. The more sensitive the bubble the greater the work that has to be done to move the liquid to its new position; hence the slower the movement.

The accuracy required in forming the surface of a sensitive bubble tube is very great. The following example gives some idea of the accuracy necessary. Suppose the tool used in forming the surface vibrates in such a manner as to superimpose a sine curve on the theoretical circle, and that the equation of this curve is $y = k \sin(2\pi x/p)$, where y is the linear departure from the true curve at any point whose distance along the curve is x, k is the maximum value of y, and p the period of the sine curve. For a displacement dx of the bubble the change of level of one end of the bubble is $(l\,dx/2R) + (y_1 - y_2)\cos\theta = h$, where l is the length of the bubble, R the radius of curvature, and θ the angle between the radius at the end of the bubble and the vertical. Thus $\theta = l/2R$, and $\cos\theta$ may be considered as unity. If h and h' are the changes in level at the two ends the corresponding angular tilt is $(h+h')/l = a$, whence

$$a = \frac{4k}{l}\sin\frac{\pi l}{p}\sin\frac{\pi dx}{p}\sin\frac{\pi(2x+l+dx)}{p} + \frac{dx}{R}.$$

The first term represents the error in sensitiveness at the point considered. This shows that in general the error is reduced with increase in l, but the term $\sin \pi l/p$ will vary from 1 to 0 as l varies. Applying this to a bubble with a mean sensitiveness of 10·3″ per millimetre (R = 20 metres) and an irregularity represented by k = ·0005 mm., p = 10 mm., length of bubble 35 mm., it is seen that the tilt necessary to move the bubble 1 mm. varies from 6·8″ to 13·8″. If the length of the bubble be increased to 40 mm. the error vanishes. If R = 100 metres (sensitiveness about 2″ per mm.) and the length of the bubble 65 mm., the tilt to move the bubble 1 mm. varies from 0·0″ to 3·9″, it will be difficult to set the bubble, and no accuracy can be obtained in reading small angles of dislevelment. The above shows that the tube must be very accurately shaped, especially if angles of dislevelment are to be read; it also shows that it is useless to attempt to calibrate a bubble tube, as the errors depend on the length of the bubble itself, and this is continually varying with changes in temperature.

§ (8) Effect of Temperature Changes.—If the temperature of the tube rises the liquid will expand and the bubble itself will contract. If the bubble become unduly short the level becomes unsatisfactory. To obviate this, sensitive bubble tubes are often made with a chamber as shown in *Fig.* 4, so that the length

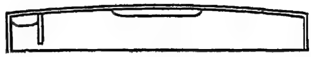


Fig. 4.

of the bubble can be adjusted by allowing more or less of the air to be caught in the chamber. If the tube be unevenly heated it will become distorted, and the same will happen if the tube be mounted in a metal mount so that it is strained by unequal expansion of the glass and metal. If the liquid be unevenly heated the warmer portion will have its density reduced and will take up more room and so cause the bubble to move irrespective of tilt. A still more important effect of unequal temperature is probably that due to the alteration of the surface tension of the liquid with change of temperature. In general the surface tension is reduced by increase of temperature, consequently the stronger surface tension at the cooler end of the bubble will cause the air bubble to move towards the warm end. It is therefore important that sensitive bubbles should be carefully shielded from becoming unevenly heated, and it is advisable to enclose the bubble tube and its mount in an outer glass tube.

§ (9) Mounting the Bubble Tube.—The bubble tube is usually fixed in its metal mount by means of plaster of Paris, but this is not suitable for sensitive bubbles. Another method is for the tube to rest on two Y-shaped supports, and be held in place by springs; the Y's and springs may be lined with cork. Another method is to fix a metal cap on each end of the tube with wax. The caps are split and one portion turned up slightly to form a spring. The caps are then secured in the metal mount by means of three screws each, one of which presses against the sprung portion of the cap. The metal tube must also be mounted so that it can be adjusted without it being strained. One end must be capable of a small turning movement, while the other is fixed by two or four antagonising screws, or by a screw working against a spring.

§ (10) Method of Viewing the Bubble.—The accuracy of setting or reading the bubble depends not only on the sensitiveness and

accuracy of the tube itself, but also on the method of viewing and illuminating the bubble. In the usual method of mounting the bubble is both lit and viewed from the top. In this case the appearance of the bubble ends varies with the direction from which the light falls, and there is also a parallax effect depending on the thickness of the glass — this being specially noticeable with a long bubble. The best method of illuminating the bubble appears to be by transmitted light, the mounting tube being cut away below as well as above, and a suitable reflector provided beneath it. If the bubble be mounted in this way the edges and ends of the bubble are better defined if the sides of the bubble tube are blackened. In many instruments, *e.g.* engineers' levels and theodolites, the bubble is so mounted that it cannot be viewed by the observer from his position at the end of the telescope; he has to move round the instrument to see the bubble, and by this movement he may tilt the whole instrument unless it stand on an exceptionally firm base. This effect can be largely overcome by viewing the bubble in a mirror, but,

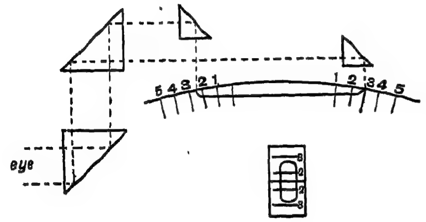


Fig. 5.

unless the mirror is at the correct inclination —and this varies according to the position of the eye,—parallax is not avoided, and the divisions of the scale at the two ends of the bubble being at different distances from the eye, they appear of different sizes. Both these effects are liable to introduce errors. A combination of reflecting prisms such as that shown in *Fig.* 5 avoids these errors, and also enables the ends of a long bubble to be viewed close together. The method of viewing the bubble shown in *Fig.* 6 (introduced by Messrs. Zeiss) entirely eliminates parallax, and enables the bubble to be set with very great accuracy. If the line *ab* of the prism is vertically over the centre line of the bubble, one half only of each end of the bubble is seen; and if the ends of the bubble are equidistant from *c*, the appearance is as shown in *Fig.* 6 (*a*). As the bubble moves to the right the two ends appear to move in opposite directions, and the appearance is as in *Fig.* 6 (*b*). The bubble is set by bringing the two ends into coincidence as shown in *Fig.* 6 (*a*). This arrangement more than doubles the accuracy of setting, as the separation of the two ends as seen in the prism is double the actual movement of the bubble, and it is much easier to bring the two ends into coincidence than to judge when the bubble is lying evenly between the marks on the tube. Another advantage of this arrange-

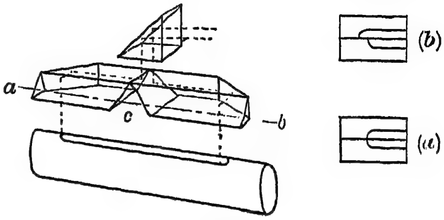


Fig. 6.

ment is that the bubble can be very easily adjusted, the final adjustment being made by shifting the prism box along the tube instead of tilting the tube itself; this is a far coarser and easier mechanical adjustment for a given angular adjustment.[1] This arrangement is probably the best so far devised for setting the bubble, but is not so suitable for use where it is required to read small angles of dislevelment. An attempt has been made to get the same facility of adjustment by marking the graduations on an outer glass tube instead of on the bubble tube itself; this method, however, increases parallax errors.

§ (11) Testing Bubble Tubes. — The "bubble tryer" shown in *Fig.* 7 is used for testing the sensitiveness and accuracy of bubble tubes. The bubble tube to be tested is placed on the Y's, the instrument is levelled laterally by the supporting screws on the right. The bar supporting the Y's can be tilted by means of the micrometer shown at the left. The dimensions of the instrument are calculated so that one revolution of the screw tilts the bubble by a definite number of seconds.

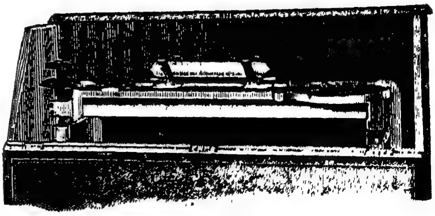


Fig. 7.

For sensitive bubbles a double lever arrangement can be used. By this means the angle necessary to move the bubble through any of the divisions of the scale can be found, hence

[1] *Fig.* 16 of article "Surveying and Survey Instruments" shows an instrument with this arrangement.

the mean angular value of a division and the constancy of this value. It will be seen from what has been said above that it is not possible to calibrate a bubble tube, as with an inaccurate tube the sensitiveness will vary with the length of the bubble itself.

REFERENCE.—*Trans. Optical Soc.* vol. xx. p. 45.

E. O. H.

SPRAYING as an alternative method to immersion for the application of silvering solutions; an economical process, used for search-light mirrors. See "Silvered Mirrors and Silvering," § (2) (iii.).

STADIA SURVEYING WORK, special instruments for. See "Surveying and Surveying Instruments," § (21) (v.).

STADIA SURVEYS, subtense methods. See "Surveying and Surveying Instruments," § (22).

STADIA SURVEYS, TACHEOMETRIC OR, general methods. See "Surveying and Surveying Instruments," §§ (9) (v.), (20).

STANDARD WAVE-LENGTHS, definition of a system of standard wave-lengths. See "Wave-lengths, The Measurement of," § (4).

Discussion as to their values. See *ibid.* § (4).

STANDARDS, OPTICAL, limits of error permissible in the manufacture of spectacle lenses. See "Lenses, The Testing of Simple," § (4).

STANDARDS OF LIGHT. See "Photometry and Illumination," § (4) *et seq.*

STARK EFFECT: an effect, discovered by Stark in 1913, in which certain spectral lines when emitted in strong electric fields are resolved into components which are polarised in different planes. The resolution is not the same for lines of the same series, but the polarisation and number of the components depend on the direction, with respect to the field, in which the observations are made. See "Spectroscopy, Modern," § (9).

STARLING-THOMPSON PHOTOMETER. See "Photometry and Illumination," § (27).

STATION POINTER, its use for fixing position of ships. See "Navigation and Navigational Instruments," §§ (14) (i.), (17) (i.).

STEFAN-BOLTZMANN LAW. See "Radiation Theory," § (5).

STEREOSCOPIC GRATICULES: a device employed in some stereoscopic range-finders to give a distance scale in the field of view. See "Range-finder, Short-base," § (13).

STEREOSCOPIC RANGE-FINDER. See "Range-finder, Short-base," §§ (2) and (7).

STIMULUS: a term used in connection with sensory phenomena to denote the external physical cause as distinct from the sensation perceived.

"STONES," a defect in glass. See "Glass," § (16) (i.).

STOPS, MEASUREMENT OF EFFECTIVE APERTURE OF. See "Camera Lenses, Testing of," § (8).

STRAIN IN TRANSPARENT MATERIALS, detection of, by polarised light. See "Polarised Light and its Applications," § (19). See also "Glass," § (19) (iv.).

Production of, in glass. See *ibid.* § (19) (i.).

"STRIAE," A DEFECT IN GLASS: veins of glass of a different refractive index from that of the rest of the glass; known also as "cords." See "Glass," § (16) (ii.).

STRING, CALCULATION OF FREQUENCIES OF VIBRATION OF. See "Sound," § (52) (i.); also "Strings, Vibrations of."

STRINGS, VIBRATIONS OF

To investigate the transverse vibrations of a tense string or wire we take the axis of x along the undisturbed position and denote by y the transverse displacement at time t of that point whose equilibrium co-ordinate is x. If T_1 be the stretching force the transverse forces on the ends of an element δx are

$$-T_1 \sin\psi, \text{ and } T_1 \sin\psi + \delta(T_1 \sin\psi), \quad (1)$$

where ψ is the inclination of the curve to the axis of x. Hence if ρ be the line density, we have

$$\rho\delta x\frac{\partial^2 y}{\partial t^2} = \delta(T_1 \sin\psi). \quad . \quad . \quad (2)$$

This equation is accurate, but if ψ is everywhere small we may neglect the changes of tension, and further write $\sin\psi = \tan\psi = \partial y/\partial x$ with sufficient approximation. Thus

$$\frac{\partial^2 y}{\partial t^2} = c^2\frac{\partial^2 y}{\partial x^2}, \quad . \quad . \quad . \quad (3)$$

where $$c = \sqrt{\left(\frac{T_1}{\rho}\right)}. \quad . \quad . \quad . \quad (4)$$

The simplest, and naturally the most important, case is where the string is uniform, and c accordingly a constant. The general solution of (3) is then

$$y = f(ct - x) + F(ct + x). \quad . \quad . \quad (5)$$

It is easily verified, in fact, that this satisfies (3), and the arbitrary functions f and F enable us to satisfy prescribed initial and terminal conditions. For instance, in the case of an

unlimited string started from rest in the form $y=\phi(x)$ the solution is evidently

$$y=\tfrac{1}{2}\{\phi(x-ct)+\phi(x+ct)\}, \quad . \quad . \quad (6)$$

since this makes

$$y=\phi(x), \quad \frac{\partial y}{\partial t}=0, \quad . \quad . \quad (7)$$

when $t=0$.

If t be increased by any arbitrary amount τ, and x by $c\tau$, the value of the first term in (5) is unaltered. Hence this term, taken alone, would represent a wave-form travelling unchanged in the direction of x-positive with the constant velocity c. Similarly, the second term in (5) represents a wave travelling with velocity c in the negative direction. From the generality of the solution (5) we infer that any initial disturbance of the string resolves itself into two waves travelling in opposite directions. In the particular case where the string is started from rest, the formula (6) shows that the two waves are identical in form, and have at corresponding points half the original amplitude.

The kinetic energy of any portion of the string is given by

$$T=\tfrac{1}{2}\int\left(\frac{\partial y}{\partial t}\right)^2\rho dx, \quad . \quad . \quad . \quad (8)$$

the integral being taken over the portion in question. Since the length of an element is altered from ∂x to sec $\psi\delta x$, the work done in stretching it is

$$T_1(\sec\psi-1)dx=\tfrac{1}{2}T_1\left(\frac{\partial y}{\partial x}\right)^2\delta x,$$

to the second order. The potential energy is therefore

$$V=\tfrac{1}{2}T_1\int\left(\frac{\partial y}{\partial x}\right)^2dx, \quad . \quad . \quad (9)$$

the integral being taken over the disturbed portion. In a "progressive" wave, *i.e.* a wave travelling in one direction only, say

$$y=f(ct-x), \quad . \quad . \quad . \quad (10)$$

we have

$$\frac{\partial y}{\partial t}=-c\frac{\partial y}{\partial x}, \quad . \quad . \quad . \quad (11)$$

and therefore $T=V$, in virtue of (4). The energy is therefore half kinetic and half potential.

When the string is limited in one or in both directions reflection will take place. For instance, if the origin be fixed we must have, in (5), $f=-F$ for all values of the variable, and therefore

$$y=F(ct+x)-F(ct-x). \quad . \quad . \quad (12)$$

If we suppose the string to lie to the right of the origin, the first term represents a wave of arbitrary form travelling towards the fixed end, whilst the second represents the reflected wave. This is anti-symmetrical to the original wave, *i.e.* it is identical in form except that it is reversed in sign, and also end for end.

In Acoustics we are concerned with strings of finite length, fixed at both ends. Supposing these to be the points $x=0$, $x=l$, the formula (12) will apply, provided

$$F(ct+l)=F(ct-l) \quad . \quad . \quad (13)$$

for all values of ct. The function F is therefore unaltered whenever the variable is increased by $2l$, and the values of y and dy/dt will therefore recur exactly whenever t increases by $2l/c$. Hence, whatever the initial circumstances, the motion of the string is strictly periodic, the period $2l/c$ being the time a wave would take to travel over twice the length of the string. In this respect strings occupy an almost unique position among vibrating systems of multiple freedom. The only other acoustical instance is that of longitudinal vibrations of uniform rods or columns of air.

To treat the question by the more general procedure of the Theory of Vibrations we inquire what modes of vibration are possible in which the motion of each particle is simple harmonic with uniform period and phase. Assuming

$$y=u\cos(\omega t+\epsilon), \quad . \quad . \quad (14)$$

where u is a function of x only, we find that (3) is satisfied, provided

$$\frac{d^2u}{dx^2}+\frac{\omega^2}{c^2}u=0, \quad . \quad . \quad . \quad (15)$$

whence

$$u=A\cos\frac{\omega x}{c}+B\sin\frac{\omega x}{c}. \quad . \quad (16)$$

Since u is to vanish for $x=0$ and $x=l$, we must have $A=0$ and $\sin(\omega l/c)=0$, or

$$\omega=\frac{s\pi c}{l}, \quad . \quad . \quad . \quad (17)$$

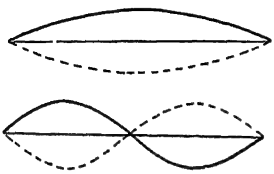
where s is integral. We infer that there is an endless series of normal modes (as they are called) of the type

$$y=B_s\sin\frac{s\pi x}{l}\cos\left(\frac{s\pi ct}{l}+\epsilon_s\right), \quad . \quad (18)$$

where the constants B_s and ϵ_s are arbitrary. The mode corresponding to $s=1$ is called "fundamental"; in it the form of the string at any instant is that of a semi-undulation of a curve of sines. In the next member ($s=2$) of the series the string forms a complete undulation, with a point of rest, or "node," at the middle point. In the mode corresponding to any other value of s there are $s+1$ nodes (counting the ends), and s intervening "loops," or points of maximum amplitude.

The frequencies ($\omega/2\pi$) of the successive modes form a harmonic series—

$$\frac{c}{2l}\times(1,\ \tfrac{1}{2},\ \tfrac{1}{3},\ \tfrac{1}{4},\ \ldots\ ).$$



This relation of frequencies has important consequences in Acoustics.[1] As is suggested by a remark already made, it is peculiar to strings and to one or two other ideally simple systems. Even in the case of a string the harmonic relation is at once violated if the density is not uniform, or if the flexural stiffness of the wire has to be taken into account.

The frequency of the fundamental mode, which determines the "pitch" of the note produced, is

$$\frac{c}{2l}=\frac{T_1^{\frac{1}{2}}}{2l\rho^{\frac{1}{2}}}. \quad . \quad . \quad . \quad (19)$$

It is lowered by increase of length or of line-density, and raised by increase of tension. These points have familiar illustrations in the structure and tuning of the pianoforte.

So far, the string has been supposed free from external force. To illustrate the case of "forced" vibrations we may imagine that a prescribed vibration

$$y=\beta\cos\omega t \quad . \quad . \quad . \quad (20)$$

is imposed at the point $x=a$. The two parts of the string are to be treated separately, since the equation (3) is violated at the point in question. The solution is

$$y_1=\frac{\sin(\omega x/c)}{\sin(\omega a/c)}\beta\cos\omega t, \quad . \quad (21)$$

for $0<x<a$, and

$$y_2=\frac{\sin\{\omega(l-x)/c\}}{\sin\{\omega(l-a)/c\}}\beta\cos\omega t. \quad . \quad (22)$$

Each of these expressions is seen to come under the form (5), whilst for $x=a$ the values of y_1 and y_2 agree with (20). The amplitude of y_1 or y_2 becomes very great whenever $\omega a/c$ or $\omega(l-a)/c$ is a multiple of π, *i.e.* when the imposed period approximates to a natural period of either of the two segments of the string. We have here an illustration of the principle of Resonance (see "Simple Harmonic Motion," Vol. I.), but it is to be remembered that when the amplitude exceeds certain limits dissipation forces become important.

It is to be remarked that the direct action of a string in starting air waves is quite insignificant. In the pianoforte, for example, the periodic pressures on the bridges near the ends of the string set the whole area of the sounding-board into vibration, and this is really the origin of the audible sound. There is, of course, a certain reaction on the string itself, but this is negligible except in so far as the loss of energy gradually brings the vibrations to an end.

H. L.

[1] See "Sound," § (52).

Submarine Range-finder (Zeiss). See "Range-finder, Short-base," § (7). See also "Periscopes."

Sub-standards of Light: electric lamps used in practical photometry instead of primary standards. See "Photometry and Illumination," § (13).

Substitution Method: application to measurement of prism angles. See "Goniometry," § (2).

Subtense Methods. See "Surveying and Surveying Instruments," § (22).

Sugar, Rotation Constants of. See "Saccharimetry," § (5).

Sugar Standard, Subsidiary, for Measuring the Rotation Constants of Sugar: a quartz plate which, when measured with a saccharimeter, gives the same reading as that of a normal solution with the same instrument. See "Saccharimetry," § (5).

Sumner Line. See "Navigation and Navigational Instruments," § (4).

Surface Tension of Glass. See "Glass," § (32).

SURVEYING AND SURVEYING INSTRUMENTS

I. Surveying in General

§ (1) Introduction.—Surveying consists in the application of physical measurement to the earth's surface, and furnishes a basis for its discussion. Land surveying may be divided roughly into three classes, cadastral, topographical, and geodetic, but there is no clear-cut division between them. Strictly speaking, a cadastral survey is concerned only with property boundaries, the topographical with natural features and physical objects, and geodetic with the shape of the earth. In any highly developed country a cadastral survey must be on a large scale, and any large scale survey is frequently called a cadastral survey. For example, the large scale Ordnance Survey maps of the United Kingdom are sometimes called cadastral maps, although they do not show property boundaries as such; they show

the physical divisions of the land such as hedges, walls, or ditches, which may or may not be the actual property boundaries. They are therefore, strictly speaking, topographical maps on an unusually large scale. On the other hand, a cadastral map is of little use if it does not include some topographical features. Generally speaking, a topographical map should show all the accidents of the surface which distinguish it from a featureless plane and which are of sufficient size to be shown legibly on the map. To satisfy these requirements on a scale larger than six inches to the mile (or about 1/10,000) would be very expensive and of little general use, and such large scale maps are only made of small areas for special engineering purposes. The smallest scale which can be considered a topographical map is about 1/500,000, maps on scales smaller than this become compilations or atlas maps. The scales most used for topographical purposes vary from 1/25,000 to 1/250,000. A geodetic survey does not in itself result in a map, but it is impossible to produce a good topographical or cadastral map of any considerable area without measurements which approach the accuracy required for geodetic purposes, and if the maps are to be on any but the smaller scales geodetic accuracy is necessary.

Land surveying is a very ancient art, and was carried out in Egypt as early as 3000 B.C., being required for laying out property boundaries on the land covered by the Nile floods, and for measuring the areas of holdings for revenue and land registration purposes.[1]

§ (2) Accuracy Required.—In deciding on the methods to be adopted in any given case it is necessary to consider on the one hand the precision of measurement and the elimination of errors of observation, and on the other the nature of the region to be surveyed and the purpose for which the results are required. Superfluous accuracy may lead to as unprofitable expenditure as an attempt to survey a large area without suitable means of controlling and eliminating the errors which must occur. For economical work there must be a proper adaptation of the means available to the end in view. The attainment of this requires a knowledge of the principles on which all survey work is based, and of the causes, and laws of accumulation, of the errors in the various processes.

The ideal accuracy is such that no error in fixing the position of any point is greater than the smallest amount that can be plotted on the largest scale map which is to be produced. While this limiting error should never be exceeded, it must be remembered that the final error in the position of a point depends not only on the immediate measurements made to it, but includes the errors of the points from which it is fixed. It is necessary therefore to provide a framework of points fixed with great accuracy, and with any errors that may appear, distributed in the best manner possible. These points are then used to control the detail of the survey, and prevent the accumulation of errors. By this means the bulk of the work can be done by rougher methods, and the most accurate, and therefore most expensive, work reduced to a minimum.

When commencing any new survey it is always economical in the long run to provide a framework, over the whole area to be surveyed, of the highest accuracy that is likely to be required in the future, and so to arrange it that it can be extended if and when the area is to be extended. If the framework is only just sufficiently accurate for a small scale topographical map, it will be useless for a larger scale, or more extensive, survey in the future. Limitations of time, or funds, available may necessitate a lower degree of accuracy, but it is generally false economy to hurry unduly or to starve the provision of this primary framework. If such an accurate framework is provided the points composing it must be marked in a permanent manner, so that they can be accurately identified at a future time; it is obvious that any failure in this respect makes the work useless for anything but the survey immediately in hand.

§ (3) Co-ordinates.—In all cases the general principles of fixing and defining the points are the same. The position of a point is defined either numerically or graphically by some system of co-ordinates, one co-ordinate being always measured vertically, *i.e.* in the direction of the force of gravity through the point. A surface which is everywhere at right angles to the vertical is a horizontal, or level, surface. The simplest example of such a surface is that of a lake or sea at rest, and the other two co-ordinates are generally measured on the surface represented by the mean level of the sea, supposed to be continued through the land. This surface constitutes the "Geoid," being the mathematical "figure of the earth" (see the article on "Gravity Survey," Vol. III.). The shape of the geoid approximates to an oblate spheroid having a semi-major axis of about 6,378,000 metres and an eccentricity of about 1/299. In surveys of large areas, such as countries or continents, it is necessary to take account of the spheroidal shape, and measurements are reduced to some selected "spheroid of reference." For smaller areas it is sufficient to consider the reference surface as a sphere whose centre lies in the earth's polar axis, and for still smaller areas the surface may be considered as a plane. The assumption that the surface is a sphere rather than a spheroid leads to the simplification that the vertical

[1] See H. G. Lyons, *Cadastral Survey of Egypt*, Cairo, 1908.

plane through a point A which also includes a point B is identical with the vertical plane through B which includes A, and that the level surfaces are all concentric spheres.

In commencing the survey of any area the reference surface is first selected, a point on this surface is then chosen as "Origin," and if the survey is to be connected to any other survey, the position of this point must be determined, either by connecting it to an existing survey, or by astronomical determination of its latitude and longitude. An azimuth must be similarly determined in order that the survey may be properly oriented. If the direction and distance from the origin of any other point is determined its position is fixed. Further points can be fixed in a similar manner, or by determining their directions from two points already fixed, or by determining their horizontal distances from two fixed points. In the latter case the height of the point must be determined separately, in the former case it is usual to measure horizontal angles only, and if this is done the height must also be determined separately. The system of horizontal co-ordinates to be adopted depends on circumstances. When an appreciable fraction of the earth's surface is to be considered the co-ordinates used are latitude and longitude.

§ (4) Latitude.—The latitude of a point is defined as the angle made by the vertical through the point with the plane of the equator. It is measured north or south from the equator. Strictly there are three different definitions of latitude—*astronomical latitude* as defined above and depending on the true vertical at the point; *geocentric latitude*, which is the angle the line from the earth's centre to the point makes with the plane of the equator; and *geographical latitude*, which is based on the supposition that the earth is a spheroid of known compression, and is the angle that the normal to this spheroid makes with the plane of the equator. It differs from the astronomical latitude only in being corrected for the local deviation of the plumb-line. It is this latitude which is used in mapping. The astronomical latitude of any point can be found direct by observation on the sun or stars, and can be found without much difficulty with an error of less than 0·3″ (corresponding to about 30 feet on the ground) using a 6-inch theodolite, and with greater accuracy by using special instruments. The astronomical latitude is not absolutely constant for any point, but varies through a range of about 0·5″ owing to the fact that the geometric axis of the earth does not coincide absolutely with the axis of rotation.

The angle between the true vertical at any point and the normal to the spheroid of reference is known as the deviation of the plumb-line, and in places where this (or rather its component in the meridian) varies the length of a degree of latitude as measured on the earth's surface will also vary irregularly; consequently astronomical latitudes are not suitable as co-ordinates for mapping purposes for large scale work. The deviation of the plumb-line may amount to over half a minute in exceptional cases,[1] and in the absence of a geodetic survey there is always an uncertainty of a few seconds. Geographical latitudes are corrected for this deviation, and therefore the length of a degree of geographical latitude does not vary irregularly. When the latitude of the origin, and the spheroid of reference to be used, have once been settled for any survey, the geographical latitudes can be calculated from the measurements and used for mapping purposes, but the resulting values will depend on the spheroid adopted and on the (necessarily unknown) deviation of the plumb-line at the origin. Consequently two surveys started independently cannot in general be expected to give the same values for the latitudes of any points common to both.

§ (5) Longitude.—The longitude of a point is the angle between the meridian plane through the point and some standard meridian plane, and is measured east or west from the standard. The meridian plane now almost universally adopted as standard is that of the transit instrument at Greenwich Observatory. The plane of the meridian may be defined in two ways, either as the plane through the earth's axis which contains the point, or as the plane parallel to the earth's axis which contains the vertical at the point. The first definition gives the *geographical longitude* and the second the *astronomical longitude*, the difference between them being the deviation of the plumb-line in the prime vertical (the plane passing through the vertical and perpendicular to the meridian). Geographical longitude is calculated in the same way as geographical latitude, but the determination of astronomical longitude is less simple. The difference in longitude between two points is a measure of the difference in the local times of the two points, one hour of mean time corresponding to 15° of longitude. Local time can be found by astronomical observations with an error of less than $\frac{1}{10}$ second without much difficulty, but to deduce the longitude Greenwich mean time must also be known. This is the most difficult part of the problem. Recent advances in wireless telegraphy have simplified this, and wireless signals can be recorded on a chronograph with a time lag of less than $\frac{1}{100}$ second, and there now appears no reason why astronomical longitudes should not be obtained in the field with the same accuracy as latitudes, and with nearly

[1] See de Graaf Hunter, *The Earth's Axis and Triangulation*, *Survey of India*, *Prof. Paper*, 16, Dehra Dun, 1918.

the same facility. In the past the only accurate method has been by exchange of telegraphic signals. This is only possible when arrangements can be made for the use of a telegraph wire between the field station and some station whose longitude is known. Transport of chronometers is not very satisfactory, as the rates of the chronometers are liable to change while they are being transported, but this has up to now been the best method available where triangulation or telegraphic determinations have not been possible. Recent improvements in chronometers, combined with the possibilities of rapid transport by aeroplane, have increased the accuracy possible by this method. A recent determination [1] of the difference in longitude between Paris and Greenwich shows the possibilities. In this Paul Ditisheim used 13 chronometers, and took in all five trips between Paris and Greenwich, giving sixty-one separate determinations. The mean for the sixty-one was 9 minutes 20·947 seconds ±·027, the probable error of one determination being 0·214 second. The mean of the French and British determinations by telegraphic methods in 1902 was 9 minutes 20·953 seconds.

Greenwich time can also be obtained by observations of the occultation of stars by the moon, or by observing or photographing the position of the moon among the stars, but such observations are neither simple nor accurate.

§ (6) "Figure of the Earth."—This is found by determining as accurately as possible the astronomical latitudes and longitudes of points connected by a geodetic survey, and the figure so found is that which makes the resulting deviations of the plumb-line a minimum for the stations considered. Naturally as geodetic surveys are extended more data become available, and fresh figures can be found which approximate more closely to the true geoid. For this reason surveys started at different times and in different countries have adopted different figures for their calculations, but these variations have little effect on the mapping of the various countries, and the labour of fresh calculations which would be necessitated by a change in the figure has not been considered to be justified by any gain in accuracy that would result. Inconvenience arises, however, when two such surveys connect up with each other,[2] and as the existing gaps in geodetic triangulation become filled up, the question of recalculation, so as to make a consistent whole, will have to be seriously considered.

[1] Monthly Notices, *Royal Astronomical Soc.* lxxx. 809.

[2] A striking example of the inconvenience caused by surveys which had not been reduced to a consistent whole, but which connected or overlapped, is given by the theatre of operations of the British Armies in France and Flanders in the late war. There were no less than five good, but conflicting, systems of triangulation in this area. See *Geographical Journal*, liii. 253.

§ (7) Rectangular Co-ordinates.—Instead of using geographical co-ordinates (latitude and longitude), it is frequently preferable to use rectangular co-ordinates, and these are invariably used in large-scale work. The position of any point on the earth's surface can be defined with reference to any other point chosen as origin by the intercept of the great circle at right angles to the meridian through the origin and passing through the point (x co-ordinate), and the distance along the meridian from the origin to the intersection of the great circle (y co-ordinate). Rectangular co-ordinates as thus defined are strictly accurate means of defining the position of points. Geographical and rectangular co-ordinates can be converted the one into the other.[3]

The rectangular co-ordinates of traverse points or trigonometrical points, whose latitude and longitude are not required, are generally calculated as follows.

A (*Fig.* 1) is the origin, AY the meridian, AX the great circle perpendicular to the meridian. Then the co-ordinates (x_1y_1) of B are given by $x_1 = \text{AB} \sin \alpha$ and $y_1 = \text{AB} \cos \alpha$, where α is the azimuth, or true bearing of the first side AB. Similarly, if $\beta = \text{ABC} - \alpha$, $x_2 = \text{BC} \sin \beta$, and $y_2 = \text{BC} \cos \beta$, the co-ordinates of C are $(x_1 + x_2)(y_1 - y_2)$. Similarly the co-ordinates of D and further points can be found. It will be noticed that in the case of a triangulation such as is shown in *Fig.* 1 the co-ordinates of C can be found either from B or A, and those of D from B or C. It is generally advisable to compute both ways so as to provide a check on the arithmetic. It must also be noticed that once the initial meridian has been left the angles β, γ, etc., are no longer the true bearings of the sides, and must never be used as such. They are commonly known as "false bearings," and differ from the true bearings by an amount known as the convergence of the meridians. This method of calculation assumes that the perpendiculars dropped from the points B, C, etc., to the axes are parallel straight lines. In reality the perpendiculars to the Y axis are portions of great circles, and therefore not parallel to each other nor to the axis of X, hence as the points get further from the Y axis the y co-ordinates as calculated become larger than their true

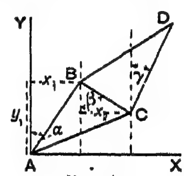


Fig. 1.

[3] See A. R. Clarke's or other book on Geodesy; or C. F. Close, *Text-book on Topographical Surveying*.

value. This error increases as the square of the distance from the meridian, and amounts to about 1 ft. per mile at a distance of 75 miles from the meridian. The errors of the x co-ordinates, and those of the y co-ordinates, due to increasing distance from the X axis, are much less serious. This method of computing, though extremely useful, must not be continued too far from the initial meridian.

§ (8) MAP PROJECTIONS.—In order to form a map the points must be plotted on a plane surface, and some orderly system of constructing the meridians and parallels must be selected for this purpose. Such an arrangement is called a map projection, although it is not, as a rule, a projection in the geometrical sense of the term. As the surface of the earth is curved it is impossible to represent it without distortion on a plane surface. For small areas, where the departure of the earth's surface from a plane is small, it does not matter much what projection is selected, and points are generally plotted by treating their (spherical) rectangular co-ordinates as plane co-ordinates. In a map covering a large area of ground the main points must be plotted either from their geographical co-ordinates, or from these reduced to plane rectangular co-ordinates, according to the projection adopted. In the case of large-scale work minor points are plotted by rectangular co-ordinates, and the consequent errors are prevented from accumulating to any appreciable extent by shifting the origin used from one principal point to another as the work proceeds. The various map projections in use are dealt with in a special article.

§ (9) METHODS OF MEASUREMENT. — The actual measurements required in the field are of two kinds:

(*a*) Length measurements, either direct by means of measuring rods, chains, tapes, etc., or indirect by the use of angular measurements. (*b*) Angle measurements. Generally horizontal and vertical angles are measured separately. The instruments most used are theodolite, compass, or plane table, for horizontal angles; theodolite or clinometer (with either a spirit-level or plumb-bob) for vertical angles; and the sextant for direct angles.

These fundamental measurements can be combined in various ways. For obtaining horizontal co-ordinates the following methods can be adopted:

(i.) *Triangulation*.—This consists of forming a network of triangles, built up one on the other, the distance between two points being first accurately measured by a base measuring apparatus. The angles of the triangles being then measured with a theodolite, the lengths of all the remaining sides can be calculated. This method is the best for providing the main framework of any extensive survey, and is the only one used for geodetic work. It should always be used where circumstances permit (see § (10)).

(ii.) *Traversing*.—Traversing consists in proceeding from point to point in straight lines. The distance between points is measured directly, and changes of direction are measured at each point by means of theodolite or compass. Traverses are used in close country, such as towns or forests, where triangulation is impossible or unduly expensive, or for railway or road location surveys where the survey of only a narrow strip of land is required. They are generally used in conjunction with a triangulation (see § (11)).

(iii.) *Detail Survey*.—Detail survey consists in the filling in of minor points of detail between the trigonometrical or traverse points already fixed. It is generally carried out by simple length and angular measurements (chain survey, etc.), or in some cases by large-scale plane-table work (see § (15)).

(iv.) *Plane-tabling*.—The plane-table consists simply of a drawing-board mounted horizontally on a portable stand. The detail is drawn straight on the board in the field. An alidade, or sight-vane, consisting of a ruler fitted with sights, enables rays to be drawn on the board in the direction of the various points. Points are fixed by interpolation or resection from known points, or by a system of graphic triangulation, or by means of directions determined graphically and distances by tacheometric means (see § (16)).

(v.) *Tacheometric or Stadia Survey*.—In this case distances are measured indirectly by the angle subtended at the instrument by a known length at the distant point. If the bearing of the point is known the co-ordinates of the point can be calculated, or its position plotted by protractor and scale (see § (20)).

(vi.) *Astronomical Observations*. — Astronomical latitudes, local time, and the direction of the true north can be obtained from observations on the sun or stars. As already explained, such observations are required for determining the position of the origin and the orientation of new surveys, and for geodetic purposes. They are also useful as a check on route traverses and rough triangulation. The determination of an azimuth is often required in many classes of survey (see § (23)).

(vii.) *Photographic Surveying*.—Photographs can be used for measuring angles if the focal length of the lens used is known. Such methods are sometimes of use in mountainous country, and the recent developments in aviation have opened up possibilities of

surveying by means of photographs taken from aircraft.

The general utility of photographic methods is, however, much debated, and for a complete topographic survey plane-table methods are generally preferable, but photographic methods may be useful in very rough country, or for location surveys in mountainous country.[1] Elaborate apparatus for plotting from photographs has been made, notably by Zeiss. A full description of the different apparatus available, together with a bibliography, is given by Max Weiss.[2]

Survey from aerial photographs suffers at present from two great disadvantages: (1) The apparent direction of gravity in an aeroplane depends on the acceleration of the aeroplane, consequently the direction of the true vertical is not known. This difficulty can be overcome if sufficient points, whose position and height are known, can be accurately identified in each photograph, but the result is that a stronger and more expensive trigonometrical control is needed than would be required for ordinary methods. (2) The difficulty in obtaining an adequate knowledge of the relief of the ground. In any case it is only the tops of walls, houses, etc., that can be fixed accurately, whereas the ground plan is usually required by surveyors. The overhang of trees and buildings creates further difficulty. These disadvantages make the method of little use in closely inhabited country. On the other hand, aerial photography proved of great use during the war to fill in fresh detail in inaccessible country, and is also of use in the survey of flat country that cannot easily be walked over, and for the rapid survey of high-water mark on the coasts or in flooded areas.[3]

The vertical co-ordinates are found in the following ways:

(*a*) *Levelling.*—Readings are taken on a vertical scale by means of a horizontal line of sight. The difference of two such readings from the same position of the instrument gives the difference of height of the two points on which the scales are supported. By a succession of such readings the heights of successive points are obtained. This is the most accurate method (see § (29)).

(*b*) *Trigonometrical Determination of Heights.*—If the distance of a point is known and the vertical angle to it measured, the height of the point above, or below, the instrument can be calculated. For distant points allowance must be made for the curvature of the earth and for atmospheric refraction. The method suffers from the uncertainty of the allowance for refraction, and the errors from this cause increase rapidly with the distance.[4]

(*c*) *Barometric Methods.*—Measurement of the atmospheric pressure by means of an aneroid, or otherwise, is the quickest method of determining heights, but as the pressure varies according to the "weather" as well as the height, this method gives only rough results in most cases.

The various methods set out above and the instruments used will now be considered in more detail.

II. Triangulation

§ (10) Triangulation.—If the length of one side, and the three angles, of a triangle are known, the lengths of the other sides can be calculated, from whence it follows that if the angles in any figure built up of triangles are measured, together with the length of any one side, the lengths of all the other sides can be calculated and the apices of the triangles fixed. The measured side is called the base, and its method of measurement is described later. The figure may consist of a chain of triangles, of quadrilaterals, of polygons, or of a network, as shown at (*a*), (*b*), (*c*), and (*d*) of *Fig.* 2.

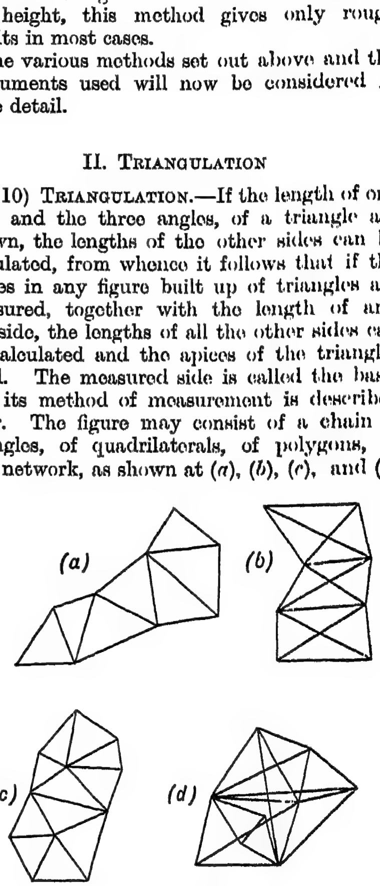


Fig. 2.

Although it is sufficient to measure two of the three angles of each triangle it is better to measure all three, as this gives a measure of the accuracy and prevents mistakes. The International Geodetic Association classes triangulation as follows: First order has a triangular error not exceeding 1″, second order not exceeding 5″, third order

[1] See F. Manek, *Österreichischer Ingenieur- und Architekten-Verein, Zeitschrift*, lxxii. 73.

[2] *Die geschichtliche Entwicklung der Photogrammetrie und die Begründung ihrer Verwendbarkeit für Mess und Konstruktionszwecke* (Stuttgart), 1913. See also R. Helbling, *Schweizerische Bauzeitung*, lxxvii. 6 and 13, and W. Sander, *Zeit. Instrumentenk.* xli. 1, 33, and 65.

[3] See *Geographical Journal*, lvi. 201 and 481; also C. G. Lewis and H. G. Salmond, *Survey of India, Prof. Paper*, 19.

[4] See article on "Trigonometrical Heights and Terrestrial Atmospheric Refraction," Vol. III.

not exceeding 15″, fourth order exceeding 15″. In a first-class extensive survey there are generally three classes of triangulation. The "primary" should, if possible, be of the first order of accuracy, or should closely approach it. The sides of the triangles would average 30 to 40 miles (up to 100 miles have been observed). All angles are measured, and a 10- or 12-in. micrometer theodolite is used. This forms a backbone, or main framework for the survey. The "secondary" triangulation is carried out with 6- or 8-in. instruments, with fewer observations, and the lengths of the sides are shorter. Errors are prevented from accumulating by adjusting the results to those of the primary work. Finally the whole country is covered by a "tertiary" network, based on the secondary work. The length of the sides depends on the scale of the map required, and for large-scale work would average about 1½ miles.

When time and funds permit, it is always advisable to provide a framework of first order triangulation. This will then serve for any survey that may be required in the future; and if accurate large-scale surveys are started in various parts of the country this framework ensures that they will join up correctly when they meet. If it is impossible to carry out such work, the main framework should be of at least sufficient accuracy to ensure that no error is introduced which will be plottable on the scale of the map which it is to control.

With any extended system of triangulation it is always advisable to measure one or more check bases, at considerable distances from the original. Bases can now be measured so easily that modern practice tends to multiply the number measured.

The errors in primary work are always adjusted by the method of least squares. The accuracy depends not only on the accuracy of measurement, but also on the "strength" of the figure. In a simple chain of triangles ((*a*), *Fig.* 2) the triangles are all independent, and the only adjustment made is by distributing the triangular error between the three angles. In other words, there are three unknown errors, and only one conditional equation that must be fulfilled by the correct angles. In the case of a chain of quadrilaterals ((*b*), *Fig.* 2), while each quadrilateral is independent, eight angles are measured, but there are four conditional equations connecting them, hence the errors can be distributed better and the figure is "stronger." The strength also depends on the figure being well conditioned; *e.g.* an equilateral triangle is stronger than one containing an acute angle. More complicated figures are still stronger, but as their strength increases so does the labour in adjusting them. Check bases provide a further correction, which must be applied throughout the chain connecting each pair of bases, it being usually considered that the measured lengths of the bases are errorless compared with those calculated through the triangulation. Similarly a chain closing on itself, or on a previously adjusted chain, so as to form a closed circuit, provides a further condition. Modern practice tends to adopt chains of quadrilaterals, forming a grid or network, with occasional more complicated figures as may be necessitated by base extensions or topographical conditions.

In first order work the signals observed to are generally lamps or heliostats, work on lamps at night giving the best results. If opaque signals are used, care must be taken to avoid a construction that presents variations in phase according to the direction from which the light falls upon it. All stations must be carefully marked and described so that they can be found again in later years. It is desirable to have concrete or brick pillars to support the instrument and the lamps. In flat or wooded country it may be necessary to erect towers to support the instrument so as to get a sufficiently clear sight.

III. Traverses

§ (11) General.—A traverse consists of a series of straight lines on the earth's surface, the lengths and bearings of which are measured. Each straight section is called a leg. The bearings of each leg may be obtained independently by compass, or the bearing of the first may be measured or assumed and the bearing of each successive leg determined by measuring the angle made with the preceding leg. Traverses may be made in many ways, and with all degrees of accuracy. The distances may be measured by pacing, by time, by counting the revolutions of a wheel run along the ground, by chain, by steel or invar tape, or by tacheometric methods, and the angles by prismatic compass, by theodolite, or by direct plotting on a plane-table. Accurate traverses may be used for the secondary framework of an extensive survey where the conditions are such as to make triangulation difficult or impossible; such conditions obtain in dense forest or flat grass country where high towers of some sort would have to be built for the stations of a triangulation, causing great expense. Modern developments in the use of steel or invar tapes for rapid and accurate measurement of length have greatly increased the accuracy, and hence the application of this type of traverse (see also § (41)).

Traverses are most widely used for roughly mapping travellers' routes, and for detail work in towns and close country.

Traverses are also used where a large-scale map of a narrow strip of country is required, such as a railway survey, and also for mining work where triangulation is impossible.

§ (12) Use of Tapes.—For town traverse work the tapes are generally stretched along the ground, the tension being applied by a spring balance, and the slope measured by the theodolite at the same time as the horizontal angles. In rough or jungle country it is convenient to use long tapes, up to 5 or 10 chains in length. In such cases it is generally best to let the tape hang free clear of the ground, supported at intervals of every chain, or some other convenient distance. For traverse purposes these intermediate supports can be wire nails driven into the sides of rods, the rods being held vertically by assistants; the supports must be aligned vertically and horizontally.

In using a spring balance for applying the tension it must be remembered that the tension indicated by the dial is that applied to the *hook* when the balance is in a vertical position with the hook downwards, *i.e.* it is the tension of the spring *minus* the weight of the drawbar and hook; the tension indicated is therefore too low, when the balance is held in any other position, by an amount $w(1-\cos a)$, where w is the weight of the drawbar and hook attached to the spring, and a the angle made by the balance with the vertical, reckoned as 0 when the hook is downwards and 180° when the hook is upwards.

In order to provide a check on the measurements it is useful to have the tape graduated on the back and front in two different units, *e.g.* feet and links, or feet and meters.

§ (13) Measurement of Angles. — The method of measuring the angles with a theodolite is similar to that employed in triangulation, but the following special considerations apply. As the bearing of any leg is burdened with the accumulated error of all the angles measured up to the point considered, the accuracy becomes of special importance irrespective of the length of the leg sighted. It is therefore of great importance, especially when using short legs, that the traverse points should be very finely marked, that both theodolite and signal should be carefully centred, and that the signal should be such that it can be accurately bisected. The longer the legs the less important this becomes.

§ (14) Computing and Plotting.—Traverses can obviously be plotted direct with a protractor and scale, but it is almost always better to calculate the co-ordinates of the points. Such co-ordinates, or their differences, are frequently known as "latitudes and departures" or "Northings, Eastings," etc.

Traverses should always be run between points already fixed, or be closed on the starting-point; this enables the total error to be determined. The errors accumulated depend on the accuracy of both the length and angular measurements, and the problem of distributing the errors found in the best manner does not lend itself to any simple solution. There are two main cases, (1) where the bearings are measured by compass, in which case the error in bearing is not accumulative, and (2) where the angles between successive legs are measured, in which case the error in bearing increases as the square root of the number of angles (assuming that all angles are read with equal accuracy). In distributing the errors it is usual to correct the angles or bearings independently of the lengths. In any closed traverse the correct sum of the angles can be calculated and the error divided equally among the measured angles. The same applies when traversing between trigonometrical points. These corrected angles are used in the calculations, and the remaining errors are distributed by corrections applied to the co-ordinates. The simplest method is to divide the total error in latitude equally among the stations, and similarly with departures. A better way (on the assumption, generally correct with careful theodolite traverses, that the length errors are large compared with the angular errors) is to use the following rule: "As the arithmetical sum of all the latitudes is to any one latitude, so is the whole error in latitude to the correction to the corresponding latitude, and so with the departures." [1]

The angular error can often be checked at intervals by observing angles to points already fixed, or by astronomical azimuths. Length errors can only be checked by closing on fixed points, or, in the case of rough exploration surveys over long distances, by astronomical observations of latitude and longitude.

IV. Detail Surveying

§ (15) General Methods.—The detail on large-scale surveys may be inserted by stadia methods (*q.v.*), by plane-table work (*q.v.*), or by chain survey. In small-scale work the plane-table is generally used. The surveyor's chain consists of links of iron or steel wire connected by rings, and with a brass handle at each end. In countries using English measures the chain is generally 66 ft. long, divided into 100 links, every tenth link being marked by a brass tab. As ten square chains are one acre, this forms a convenient decimal system for the measurement of areas; 100 ft.

[1] See J. B. Johnson, *Theory and Practice of Surveying*, 17th ed. p. 235.

and 20 metre chains are also used. The area to be surveyed can be split up into triangles (generally starting from a triangle of the tertiary triangulation) the sides of which are measured. The position of detail points are fixed by the distance along the line chained and the distance from the nearest point on the line measured (at right angles) by means of a graduated rod or tape. Such distances are called "offsets." Offsets should be kept short, or the error in estimating the right angle will become too great. Cross staves or optical squares can be used to give the right angle if long offsets are necessary. The simplest form of optical square consists of a pentagonal prism with reflection at two faces including an angle of 45°. It is more accurate, and now more usual, to use a graduated steel tape instead of a chain. Allowance for slope is made by estimation, or by measuring the slope by clinometer and calculating the correction. The field work is booked in a field book, all distances being booked as total lengths from the start of each chained line. Although the chaining of the three sides of a triangle, or the four sides and one diagonal of a quadrilateral is sufficient to determine it, it is advisable to chain also a tie line, *i.e.* a line from the apex of the triangle to a known point on the base, or the second diagonal of a quadrilateral. This gives a check, and tests the accuracy of the work when plotted.

V. Plane-tabling and Plane-tables

§ (16) Plane-tabling.—The plane-table consists essentially of a drawing-board mounted horizontally on a portable tripod, and capable of being adjusted in azimuth. Accessories used with it are a ruler provided with sights (either open or telescopic), a compass, and a clinometer. The sighted ruler is called the "alidade." The plane-table is used for topographic survey, the detail being drawn in direct on the paper, and as a rule on the scale required for the finished map. Such scale may be anything from about 1/250,000 upwards. It is not advisable to work to a larger scale than that of the final map as this entails a larger number of field sheets, limits unnecessarily the number of trigonometrical points that can be plotted on the sheet, and leads to an unnecessary amount of detail being plotted, and hence to a waste of time. The only advantage is that a lower standard of draughtsmanship can be used. If a larger scale map is likely to be required in the near future it should, however, be considered whether the larger scale should not be adopted at once for the field work.

The general principle of the work is as follows. A certain number of points are fixed first by triangulation, and carefully plotted to scale on the field sheet. The distance between such points depends on the scale, a convenient rule being that they should be about 4 in. apart on the field sheet. If the plane-table be now set up and levelled at any fixed point A on the ground, the sight rule laid along the line AB (where B is another fixed point visible from A), and the table turned till the distant station B is intersected by the sights, the table is said to be oriented, and if the sight rule be laid along the line joining A to any other fixed point, that point should be intersected by the sights. Similarly the line giving the direction of any other point from A can be drawn in; two such lines to any point from two different settings up of the table fix its position, and any subsequent lines serve as a check. Stations so fixed and checked are termed intersected points, and can be used in fixing further points. If the distance of any point be known, or estimated, this distance can be laid off along the line and the point fixed. Detail in the neighbourhood of any fixed and occupied point can thus be drawn in. The accuracy of measurement or estimation required depends on the scale, as nothing is gained by an accuracy of measurement greater than the accuracy of plotting. It is always desirable, when the table is first oriented, to mark the magnetic north by means of the compass, as the plane-table can then always be oriented approximately by compass.

If the table be set up at any unknown point from which three fixed points A, B, and C can be seen, the position can be determined by "interpolation" or "resection."[1] There are various ways of doing this, but the simplest is as follows. The table is approximately oriented by compass, and rays are drawn from the three points; if the orientation is correct these rays will meet in a point, which is the required position. If (as is probable) they do not meet they will form a small triangle, known as the triangle of error, and the true position can be determined by the following rules: (1) If the triangle of error lies within the triangle formed by the three points, the position is within the triangle of error; (2) If the triangle of error falls outside the triangle formed by the three points, the position is either to the left of all the rays when facing the fixed points, or to the right of all; (3) The distance of the point from any ray is proportional to the length of the ray. From these rules the position of the point is estimated and marked, the sight rule is placed along the line joining this position to that of the most distant of the fixed points, and directed on

[1] In America the term resection is also applied where one ray has been drawn to the point from a previously occupied point, and the fixing is completed by rays drawn at the unknown point from fixed points.

the point by revolving the table, and the orientation is then checked on the other points, or on any further points visible. If there is still an error the process must be repeated. This method fails if the position of the station lies on a circle passing through the points A, B, and C, and is inaccurate if it lie near such a position.

In small-scale work the stations at which the table is set up are fixed by intersection or resection, and the detail sketched in round them. Fixings should always be made from trigonometrical points if possible, and if not, from points that have been carefully intersected or resected and checked. By this means errors are prevented from accumulating.

(i.) *A Plane-table Traverse.*—The plane-table is not suitable for use in forest country, where sights to distant objects can seldom be obtained for fixing the positions, but it may be necessary, even in generally open country, to follow a stream or path where ordinary fixings cannot be made; in such a case it may be necessary to carry out a plane-table traverse. The plane-table is set up and oriented at the start of the traverse, and a ray drawn in the direction of the first leg (all legs should be as long as possible); the distance to the first station is measured and its position marked. The table is then set up at this station and oriented by the back ray, and the direction of the second leg drawn, and so on. As soon as a position is reached which can be fixed by ordinary methods the traverse should be adjusted as follows: if 1, 2, 3, . . . 6 are the stations of the traverse and the true position of 6 be found to be f, draw lines $2b$, $3c$, . . . through 2, 3, . . . parallel to $6f$ so that their lengths bear the same proportion to $6f$ as the lengths 1-2, 1-3, . . . bear to 1-6, then b, c, d, . . . are the adjusted positions of the traverse points. If the legs are necessarily very short and the traverse long, errors in direction may accumulate rapidly, and it may be better to make a compass traverse and plot it afterwards (see "Traverses," *ante* § (11)).

(ii.) *Graphic Triangulation.*—A graphic triangulation can be carried out if necessary, but this should be avoided if possible except for such special purposes as a preliminary reconnaissance for theodolite triangulation, etc. For such purposes a telescopic alidade should be used. Special care should be taken to get well conditioned triangles, and if possible the graphic triangulation should be tied in to a theodolite triangulation and adjusted to it.

The above method of using the plane-table is the usual British[1] practice.

[1] C. F. Close, *Text-book of Topographical and Geographical Surveying* (2nd ed., 1913); *Report on the Topographical Survey of the Orange Free State*, 1905–1911, published by the War Office, 1913.

(iii.) *American and Continental Practice.*—This has been to use the plane-table for larger scale work, and to provide it with more elaborate adjustments for levelling and orienting. Telescopic alidades are used, and detail fixed by means of stadia readings, taken either from a separate instrument or by means of stadia hairs in the alidade telescope. This method[2] gives the position and height of all points at which the stadia rod is held up to a distance of about ¼ mile from each station occupied by the plane-table.

§ (17) PLANE-TABLES AND ACCESSORIES.—The British Army pattern of plane-table is the simplest form and consists of a drawing-board 18 × 24 in. attachable to a simple tripod by a screw. The levelling is done entirely by suitably placing the legs. When

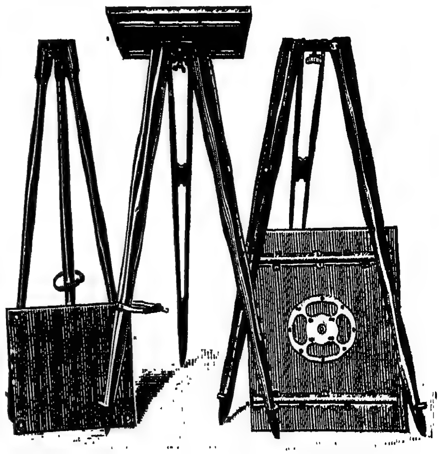


FIG. 3.

the screw holding the top is slacked off, the top can be turned horizontally. *Fig.* 3 shows three simple patterns of plane-tables: the centre one has a "Johnson" ball-and-socket levelling head. In this case the upper clamping screw clamps the top so that it cannot tilt but can rotate in azimuth until the lower screw is tightened. Smaller patterns and stands with folding legs are also used, but folding legs, though more portable, are apt to be less rigid. It is convenient on steep ground to have one leg whose length can be adjusted. The paper is usually mounted on linen, and then damped and stretched over the board, the edges turned over and glued to the back. When dry this forms a good smooth working surface which does not cockle, but suffers from the disadvantage that the

[2] J. B. Johnson, *Theory and Practice of Surveying* (17th ed., 1914); *Coast and Geodetic Survey Report*, 1905, Appendix 7.

paper distorts considerably on being removed, and it is impossible to remount the paper when once it has been taken from the board. As a result the trigonometrical points can only be plotted after the paper is mounted, and if the field sheet has to be sent for further points to be plotted, or for revision or amplification, the whole plane-table top must be sent, and no such work can usefully be done when the paper has once been dismounted. Even if the paper is kept mounted a certain amount of distortion occurs, as the table top itself expands and contracts with varying moisture, and does so unevenly, expanding more across than along the grain of the wood. A method tried with some success in India consists in pasting the field sheet (preferably of Bristol board) firmly along one long edge of the plane-table by means of a strip of cloth. The other three sides are cut so as to leave a ½-in. margin of plane-table top. Strips of cloth are then pasted along the upper surface of the field sheet and the under surface of the plane-table. The cloth is pulled tight but the paper is not wetted. If the wood now shrinks the cloth will get slack but the sheet will not cockle.

In any case of paper, Bristol board, or paper mounted on wood, more or less expansion, etc., will take place; and although uniform expansion or contraction can be remedied by photography for final reproduction, it complicates the comparison of edges of adjacent sheets and the plotting of trigonometrical points, and if the expansion is not uniform it necessitates re-drawing for the final map. A method [1] which appears to overcome all these difficulties is to mount the paper on linen, and then paste or glue this down to an aluminium sheet. The aluminium is grained to take the glue and is the same size as the top of the table; the linen extends about 3 in. over the sides and ends of the aluminium sheet; these extensions provide overhanging flaps which are stiffened by aluminium strips. The flaps are turned under the table top, and spring hooks engage in holes in the strips and keep the whole firmly attached to the board (see *Fig.* 4). The plane-table top itself need only be a framework, and the whole arrangement is lighter than the ordinary pattern. The field sheets form a permanent record which can be attached to or detached from the plane-table as often as may be required.

American surveyors usually clip the paper to the board, or clamp it down by countersunk screws; the use of celluloid sheets with such clamping screws is satisfactory, but even this is liable to cockle.

For large-scale work, particularly when a telescopic alidade is used, it is desirable to have arrangements for more accurate levelling, and both ball and socket heads and levelling screws are used for this purpose.

FIG. 4.

A slow-motion traversing screw is also desirable. These attachments add considerably to the weight of the outfit. *Fig.* 5 illustrates such an outfit.

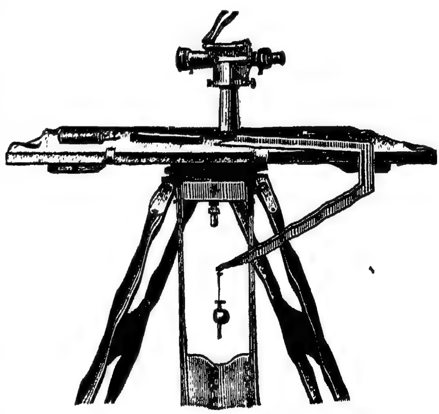


FIG. 5.

§ (18) ALIDADE.—The alidade or sight-rule should be about as long as the short side of the table. The simplest form consists of a wood rule with folding sights, the back sight having a vertical slit and the fore sight carrying a stretched vertical thread. The precise

[1] H. St. J. L. Winterbotham, *Royal Engineers' Journal*, 1919, xxx. 233.

parallelism of the sight line and the edge of the rule is of no importance provided the same edge of the rule is always used. It is a great convenience, and with telescopic alidades almost essential, to have the actual edge used for ruling on an arm moving parallel to the rule; this makes it unnecessary to pivot the alidade on the point, the arm being moved up as required after the alignment has been made. For ordinary small-scale topographic work based on an adequate triangulation a simple sight-rule is the best.

The telescope, when used, should be similar to that on small theodolites, and be provided with stadia hairs, and with a vertical arc

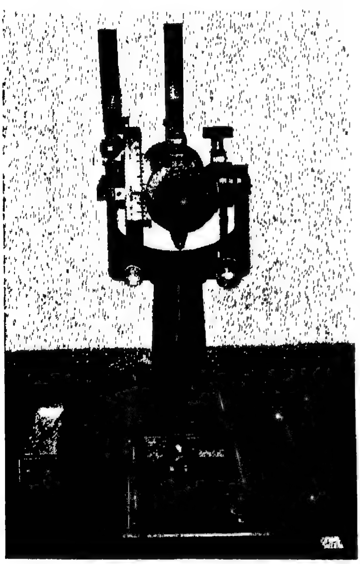


FIG. 6.

and spirit-level for reading angles of elevation and depression. No iron or steel should enter into any part of the alidade or plane-table. *Fig.* 6 shows a telescopic alidade with parallel motion and "Beaman Arc" (see also § (21), (v.)).

§ (19) ACCESSORIES.—The most suitable type of compass is a simple trough compass, 4 to 6 in. long, preferably fitted with a lid to protect the glass, and with a catch which lifts the needle from its pivot when the lid is closed (see *Fig.* 6).

For small-scale work no plumb-bob or special centring device is needed; a foot or two in the position of setting up will make no difference, but in large-scale work the table should be set up so that the plot of the point is vertically over the station mark. A plumbing bar is useful for this purpose. The illustration (*Fig.* 5) is self-explanatory.

VI. TACHEOMETRIC OR STADIA SURVEYS

§ (20) GENERAL METHODS. — Tacheometry means literally "quick surveying," and the general principle is illustrated in *Fig.* 7.

O is the observer, AB a rod of known length, C being the mid-point of AB, and OC perpendicular to AB. Then if α is the angle AOB,

$$OC = \frac{AB}{2}\cot\frac{\alpha}{2}.$$

If the rod be not perpendicular to OC, but takes a position A′B′, making an angle β with AB, then the formula becomes

$$OC = \frac{AB}{2}\cot\frac{\alpha}{2}\cos\beta.$$

If OC is inclined to the horizontal by an angle γ, then the horizontal distance of C from O is OC $\cos\gamma$, and the height of C above O is OC $\sin\gamma$. If the direction of OC is known, then the position of C with respect to O can be determined at once from measurements taken at O. This method is in many cases more rapid than chaining and levelling, and especially useful where it is difficult to measure the distance OC directly.

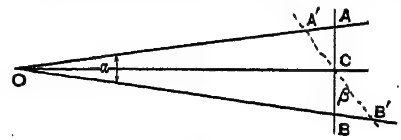


FIG. 7.

The practical methods of applying this fall into two cases: (1) where α is a fixed angle and AB is a graduated rod, the intercept AB being observed from O; this is generally called stadia surveying; (2) where AB is a known length and the angle α is measured; this is generally called the subtense method, though this term is used differently by different writers. In subtense work the angle α may either be measured by the arc of a theodolite, or by a micrometer eyepiece in the telescope.

§ (21) STADIA WORK.—The angle α is determined by two cross hairs in the focus of the telescope, and the portion of the graduated staff (stadia rod) intercepted by these two cross hairs is read off. The stadia rod is either an ordinary levelling rod, or a similar rod with rather larger divisions for use at long ranges. Stadia methods are chiefly of use in filling in the detail in large-scale topographic work.

(i.) *General Methods.*—In *Fig.* 8 O is the object-glass of an ordinary telescope of focal length f, S is the rod, the image of AB (length l) being formed at ab where a and b are the stadia hairs of the telescope at a fixed distance p apart. V is the vertical axis of the telescope.

c is the distance VO, e the distance FS, where F is the first focal point of the object-glass. A ray of light from A which passes through F will meet the object-glass in a' and be refracted parallel to the optical axis to the cross hair a, similarly with a ray from B. The figure shows at once that $e/l=f/p$, and $R=e+f+c=(l/p)f+f+c$, irrespective of the distance the cross hairs ab have been racked out in order to focus the staff. Here p, f, and c are all constants, so we can write simply $R=kl+m$, where k and m are constants. The

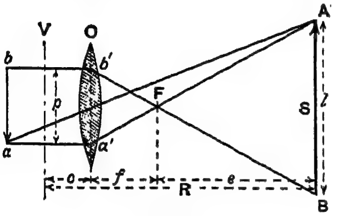


FIG. 8.

cross hairs are usually placed so that $k=100$, but sometimes $k=66$ (in which case l in feet gives e in chains), or $k=200$ (for long-distance work). The value of $m=f+c$ can be obtained with sufficient accuracy by measuring the distance (f) from the centre of the object-glass to the diaphragm when the telescope is focussed on a distant object, and from the centre of the object-glass to the vertical axis (c).

It is usual in stadia work to hold the rod vertical, as being more convenient than holding it horizontal and perpendicular to the line of sight, or perpendicular to the line of sight in a vertical plane. This results in a rather more complicated formula. In *Fig.* 9

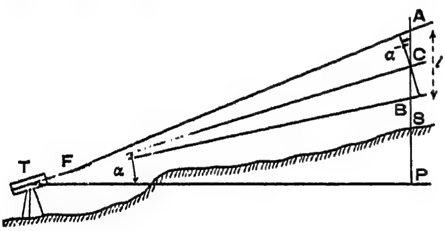


FIG. 9.

T is the instrument, S the foot of the rod, C the mid-point between A and B (the points read on the staff), P the projection of S on the horizontal plane through T; then the distance TC is $kl\cos\alpha+m$, the horizontal distance TP is $kl\cos^2\alpha+m\cos\alpha$, and the difference in height PC is $kl\cos\alpha\sin\alpha+m\sin\alpha$. If the difference in height of the ground at S and T is required, the height of the instrument must be added to PC and the distance CS subtracted.

When the stadia method is used for plotting detail and fixing points which will not be used for fixing further points by calculation, there is no object in measuring the distance to a greater degree of accuracy than it is possible to plot. A point cannot be plotted with a greater accuracy than about $\frac{1}{100}$ in., consequently on a scale of $\frac{1}{2500}$ an error up to 2 feet in the length is of no importance, and with smaller scales still larger errors have no effect. For such purposes, therefore, the constant m can be neglected. In the same way the reduction to the horizontal can always be neglected when the slope is less than 2° and generally up to 6°. If stadia traverses are being run, and the co-ordinates of the stations calculated, then greater accuracy is required, and m should be allowed for. Tables of $\cos^2\alpha$ and $\cos\alpha\sin\alpha$ are given in most text-books dealing with stadia surveying.

It is always desirable to check the interval between the stadia hairs, and to check whether they are equidistant from the centre cross hair, and for accurate work this should always be done. However accurately the makers have spaced the cross hairs, the value of k will vary owing to atmospheric refraction. The rays of light from the staff to the telescope are not straight lines, but are almost invariably concave downwards; and the amount of this bending not only varies with the state of the atmosphere, but also depends on the height of the ray above the ground, consequently both readings of the staff will be too low, and the lower reading will have a greater error than the upper. The result is that the value of k depends on the atmospheric conditions, so it is desirable to determine k for the average conditions actually met with. The determination can be made by setting up the rod at known distances from the instrument, when each reading will give a value of k if the value of m has already been determined by measurement. The conditions should be as nearly as possible those that will be met with in the work in hand, and specially favourable conditions should not be chosen; the mean value of k found should be adopted in calculating the work. An incorrect value of k will affect the scale of the resulting map, so if traverses are being run between trigonometrical points, a comparison of the distance between two such points as determined by the traverse with that determined by the triangulation will give a value for k which is a good average value for the conditions of the work in question. If the value of k to be used is not a round number it is best to construct a table connecting the stadia readings with distance, and if this be done the correction m can be allowed for in the table. Some makers provide instruments with adjustable stadia hairs, but these are not to be recommended as they are more

liable to accidental change in the value of k, and the saving in labour in reducing the results is very small.

(ii.) *Anallatic Telescope.*—With a view to obviating the necessity for the constant m Porro proposed in 1823 the use of a second lens (termed the anallatic lens) arranged as in *Fig.* 10, where C is the second (convex)

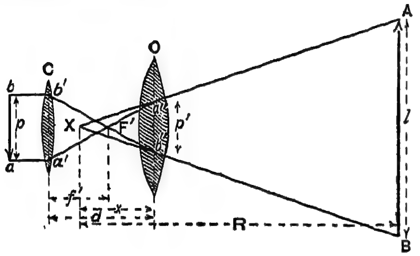


FIG. 10.

lens, whose first focal point is at F′. A ray of light from A, striking O at a'' and then passing through F′, will be refracted at a' to a, aa' being parallel to the axis. The distance d is fixed, and focussing is carried out by moving the cross hairs ab. This being the case $a''b''=p'$ is constant, and if the lines Aa'', Bb'' produced meet in X it is seen from the figure that $R/l=x/p'$=constant, and that $R=lk$. Also $p'=(d-f'/f')p$, and as X is the virtual image of F′,

$$\frac{1}{d-f'}-\frac{1}{x}=\frac{1}{f'},\quad \text{or } x=\frac{f(d-f')}{f+f'-d},$$

hence

$$R=\frac{f(d-f')}{f+f'-d}\cdot\frac{l}{p'}=\frac{ff'}{f+f'-d}\cdot\frac{l}{p}.$$

If the values of f, f', and d be so chosen that X falls on the vertical axis of the instrument, the constant m of the original formula is eliminated. The formula also shows that if the cross hairs are not quite correctly spaced, a small adjustment in the distance apart of the two lenses will enable the constant to be corrected; while this adjustment will displace the point X from the vertical axis, the error introduced will be so small as to be negligible.[1]

This construction involves not only an extra lens, but either a more bulky telescope or a falling off in its optical qualities, and it is generally considered that these disadvantages are not compensated for by the fact that no additive constant is required. It has already been pointed out that for accurate work it is desirable to use tables connecting the distance with the stadia readings, and if this be done the constant entails no inconvenience, while for rough work it can be ignored.

(iii.) *Stadia Work with Internal Focussing Telescope.*—A telescope with a negative internal focussing lens has many advantages (see § (30), (i.)), and it must be considered how far this is suitable for stadia work. Such a telescope is illustrated diagrammatically in *Fig.* 11. O and N are the object-glass and −ve lens, with their first focal points at F and F′, and focal lengths f and f'. In the absence of the lens N, an image ab of the staff would be formed at I, distant f_2 from O, where $1/L+1/f_2=1/f$, and the size of the image p would be such that $L=(f/p)l+f$. If the lens N be now inserted, the ray which previously passed along AF$a'''a$ will be refracted by N at a'' towards a', and F′$a''a'$ will be a straight line; also a ray from A

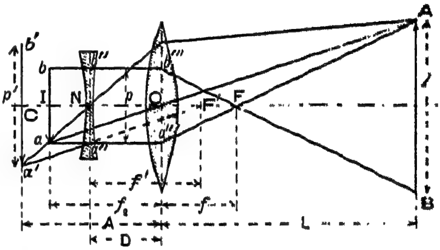


FIG. 11.

passing through the centre of N will not be refracted by N but will pass through a and a', and an image will be formed $a'Cb'$.

Then, from the figure,

$$\frac{p'}{p}=\frac{a'b'}{a''b''}=\frac{CF'}{NF'}=\frac{f'+A-D}{f'},$$

where A is the distance of the cross hairs from the object-glass, hence

$$L=\frac{f(f'+A-D)}{f'p'}l+f.$$

Here A is a constant, and D increases as L decreases. Call D_∞ the value of D when L is very large, so that $D=D_\infty+d$, then

$$L=\frac{f(f'+A-D_\infty)}{f'p'}l-\frac{fld}{f'p'}+f.$$

As D_∞ is constant put

$$\frac{f(f'+A-D_\infty)}{f'p'}=k,$$

whence

$$L=kl-\frac{kd}{f'+A-D_\infty}\cdot l+f.$$

This shows that L varies with d as well as with l, and that no simple accurate formula of the form $R=kl+m$ holds. It is simplest

[1] J. Porro, *La Tachéométrie*, Paris, 1858; see also *Zeits. Instrumentenk.*, 1882, pp. 117 and 157, and 1885, p. 413; and A. E. Young, *Inst. Civ. Eng. Proc.* cxxxix. Pt. 1.

to consider k as a constant and ascertain how m varies.

We have
$$\frac{A-D}{f_2-D}=\frac{p'}{p}=\frac{f'+A-D}{f'},$$

hence
$$f_2=\frac{-D^2+DA+f'A}{f'+A-D}.$$

Put $D_\infty=nA$, $(n<1)$, then

$$f=\frac{A^2n(1-n)+f'A}{f'+A(1-n)},\quad f'=\frac{A(1-n)(f-nA)}{A-f},$$

$$f'+A-D_\infty=\frac{A^2(1-n)^2}{A-f}.$$

$$f_2=\frac{-d^2(A-f)+Ad(1-2n)(A-f)+A^2f(1-n)^2}{A^2(1-n)^2-d(A-f)},$$

$$L=f.\frac{-d^2(A-f)+Ad(1-2n)(A-f)+A^2(1-n)^2f}{-d^2(A-f)+d\{A^2(1-2n)+f(2nA-f)\}},$$

$$kl=\frac{(L-f)A^2(1-n)^2}{A^2(1-n)^2-d(A-f)}$$
$$=\frac{f^2A^2(1-n)^2}{-d^2(A-f)+d\{A^2(1-2n)+f(2nA-f)\}}.$$

The error in assuming that $L=kl$ is E, where $E=L-kl$, hence
$$E=\frac{f\{A(1-2n)-d\}}{A(1-2n)+f-d}.$$

If we put $n=\frac{1}{2}$ as indicated in British Patent 25622 of 1910, then E reduces to $-df/(f-d)$, which is zero when the telescope is at stellar focus. The apex of the tacheometric angle (point X of *Fig.* 10), from which the distance given by $L=kl$ is measured, is therefore in the object-glass for long distances, and moves slowly backwards as the range diminishes. Taking a telescope with $A=1$ ft., $n=\frac{1}{2}$, $f=0{\cdot}9$ ft., then $f'=2$ ft., and E does not reach 0·1 ft. until L is reduced to about 25 ft., hence with such a telescope the error caused by taking $R=kl$ (R is measured from the vertical axis) will be about 6 in. at long ranges and reducing to about 5 in. at 25 ft. For detail work this is quite negligible; even for accurate traverse work it is generally negligible, as the errors due to reading the staff, and refraction errors, will be greater than this correction unless the lengths of the sights average less than, say, 100 ft. In such a case the correction to be applied is best obtained by direct trial.

If n be increased the apex of the tacheometric angle is thrown back, and the error in assuming that the telescope is anallatic is still further reduced, but for this a rather short-focus negative lens is necessary, and the optical qualities of the telescope suffer. (See the *Transactions of the Optical Society*, xxii. 20.)

(iv.) *Accuracy of Stadia Work.*—With suitable instruments and rods an accuracy of from 1/200 to 1/500 is obtained in individual sights. This error is not systematic, and, therefore, a traverse of, say, 16 legs should have an error of from 1/800 to 1/2000. About $\frac{1}{4}$ mile is the limiting length of sight for ordinary work, but up to $\frac{1}{2}$ mile can be used. An ordinary telescope, such as is used on a 5-in. theodolite, will read ordinary levelling rods up to about 400 ft.; for longer distances the $\frac{1}{10}$-ft. divisions only can be read, and for over 1000 ft. the rods should be graduated in yards or metres and tenths, or in links. The error due to refraction increases as the square of the distance, consequently long sights should not be used when the line of sight is close to the ground; such conditions are favourable for rapid and accurate work with the chain or graduated tape, but unfavourable for stadia work. On the other hand, uneven, steep, or broken ground is more favourable to stadia work, which, under such conditions, will give quicker and more accurate results than chaining.

If the stadia rod be held horizontally the refraction errors are eliminated, so far as they affect the horizontal distances, but it is necessary to provide a stand to support the rod, and to fit it with a sight to ensure its being held perpendicular to the line of sight. This is generally not considered worth while.

(v.) *Special Instruments for Stadia Work.*—Various special instruments have been made for stadia work, but they have little advantage over an ordinary theodolite, provided the latter has a suitable telescope. Greater magnification and resolution are required to read a graduated staff then to bisect a suitable target with equal accuracy.

An adjunct to enable the reduction to the horizontal etc., to be made rapidly, designed by Mr. W. M. Beaman, U.S. Geological Survey, is useful when a telescopic alidade is used for stadia working with a plane-table (see § (18)). In *Fig.* 9, if the angle of elevation α is such that $CP=nl$, and ignoring the additive constant m, we have $nl=kl\cos\alpha\sin\alpha=\frac{1}{2}kl\sin 2\alpha$, or if $k=100$, $\sin 2\alpha=\cdot 02n$. The values of α corresponding to $n=1, 2, 3\ldots$ are marked on a supplementary scale (known as the V scale) on the vertical arc. In use the telescope is set so that the index corresponds with a mark (n) on the V scale, and a full stadia interval can be read on the rod. If the centre cross hair then read $c(=CS)$, and the difference between the stadia readings be l, we have the height of S above the instrument is PS, and $PS=PC-SC=nl-c$. A second (or H) scale is graduated on the vertical arc so that any reading m on this scale indicates that the horizontal distance is m per cent less than the distance given by kl. From *Fig.* 9, $TP=kl\cos^2\alpha=kl(1-m/100)$, hence $\sin^2\alpha=m/k$. In practice the V scale is set exactly to a mark, in which case the H scale will generally not be on a mark, but by interpolation the percentage to be deducted on account of the slope can be estimated with sufficient accuracy. The alidade shown in *Fig.* 6 is fitted with a "Beaman" arc.

§ (22) SUBTENSE METHODS.—A "Subtense" bar, 10 or 20 ft. long, has been used with good results with an ordinary theodolite. The 20-ft. bar has discs 12 in. in diameter at its ends,

and is supported horizontally on a stand. A sight attached to the bar enables it to be set perpendicular to the line of sight. The angle is observed by the repetition method on the horizontal arc of the theodolite, some ten repetitions being taken. A 20-ft. bar can be used up to three miles with reasonable accuracy. The angle can also be measured very conveniently by an eyepiece micrometer, a number of readings can be taken very rapidly, and the mean of, say, 10 should give a good result. A table should be prepared giving the distance in terms of the micrometer reading, and this table should be prepared from the results of actual trial. If D is the distance from the vertical axis of the instrument to the bar, L the length of the bar, M the micrometer reading, and f the focal length of the object-glass, and c the distance of the object-glass in front of the vertical axis, we have $D = A/M + B$, where A and B are constants, as in stadia working. $B = f + c$ and A depends on the graduation of the micrometer and on the focal length of the object-glass. This method is neither so convenient nor so rapid as stadia working, but as a suitable target can be bisected with greater accuracy than a graduated staff can be read, it may be more accurate, and it can be used for longer distances.

Another method, sometimes used, is to hold the rod vertical and read the angles of elevation or depression to the two targets. In this case (see *Fig.* 12), if the angles AOP, BOP are θ and ϕ respectively, $OP = AB/(\tan\theta - \tan\phi)$

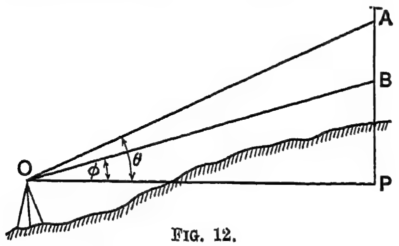


FIG. 12.

and $BP = OP \tan\phi$. If a suitably graduated horizontal scale be fitted on the theodolite at a fixed distance below the horizontal axis, the values of $\tan\theta$ and $\tan\phi$ can be read directly by means of a microscope attached to the telescope and perpendicular to it, the divisions of the scale are uniform, and the fractions of a division can be read on a micrometer which moves the scale longitudinally. An alternative arrangement is to elevate the telescope by a screw acting directly on an arm attached to the telescope so that the travel of the screw is proportional to the tangent of the angle of elevation. Various other arrangements have been made for giving more or less direct readings of the distance and height of the rod, but they are little used; their advantages for the special work are small, and they reduce the convenience of the instrument for use as an ordinary theodolite.

A variation of the subtense method is sometimes useful for executing traverses where triangulation or tape traverses are impossible. In this case signals are put up by an advanced party, and the angle subtended by these is measured as in the case of ordinary subtense work. The signals must be carefully plumbed, or else arranged so that their intersection with the ground can be observed. The distance between the signals should subtend some eight to twelve minutes, and they should be placed in a line perpendicular to the line of sight. A third signal for the actual traverse station is sometimes desirable in addition. When the observer reaches the position he must measure the distance between the signals with a tape.

VII. Astronomical Determination

§ (23) General.[1]—Astronomical determinations are used in survey work for determining latitude, local time (hence, if Greenwich time be known, longitude), and azimuth, or true north. For survey work on land the instrument usually employed is the transit theodolite. Where possible observations are made on stars, but the sun can be used. It must be remembered that latitudes, etc., thus obtained are *astronomical* latitudes and differ from the geographical values by an amount depending on the deviation of the plumb-line. Astronomical observations of position give, therefore, no check on traverse work, etc., unless the error of the traverse is likely to be greater than that due to the deviation of the plumb-line, as well as those due to the astronomical observations themselves.

§ (24) Latitude.—The usual method of determining latitude for survey purposes is by "circum-meridian altitudes." The altitude of a star when crossing the meridian, together with the known declination of the star, gives the latitude. Single observations with a theodolite suffer from errors of collimation, etc. (see article on "Theodolite"), and are of little value; hence pairs should always be taken. The usual field practice is to start observing eight to ten minutes before the star reaches the meridian, and to take as many observations as possible before transit, and an equal number afterwards. The observed altitude and time of each observation are booked, and the "face" of the telescope is changed after the first, third, fifth, etc., observation. The interval in time between each observation, and the moment of transit, being known, a correction can be

[1] For astronomical determinations for geodetic purposes see article on "Gravity Survey," Vol. III.

applied to the observed altitude to "reduce" it to the true meridian altitude. A correction must also be made for atmospheric refraction.[1] This correction is always rather uncertain, but can be largely eliminated by observing one star that transits to the north, and another that transits to an approximately equal altitude to the south; the mean of the pair will be largely free from refraction errors. As refraction is more variable for low than high altitudes, it is inadvisable to observe stars at a less altitude than 40°. Using a 6-in. micrometer theodolite, and taking the mean of several pairs of north and south stars, there should be no difficulty in obtaining a probable error of $\frac{1}{4}$ to $\frac{1}{3}$ second in the latitude.

§ (25) Observations for Time.—Local time is obtained directly from the moment any astronomical body, whose right ascension is known, crosses the meridian, and this is the method employed in fixed observatories and for geodetic purposes in the field; it is not, however, generally possible for survey purposes. If the latitude be known, and the altitude of a star be observed when it is some distance from the meridian, the local time of the observation can be calculated. The star is moving most rapidly in altitude when on the prime vertical (*i.e.* due east or west of the observer), and this is the most favourable position for observation. The latitude need not be known with any great accuracy provided the star be observed near the prime vertical. Observations should be paired, as in the case of latitude observations, both for change of face of the theodolite and for east and west stars. With a 6-in. theodolite there should be no difficulty in getting a probable error of $\frac{1}{4}$ to $\frac{1}{3}$ second in the result.

Other methods for determining latitude and time are:

§ (26) Equal Altitude Method.—This has come into favour recently. The simplest form of the method is to observe the moment a star reaches a given altitude before crossing the meridian, and the moment it reaches the same altitude after crossing. Half-way between these two times the star must be on the meridian, hence the time of transit is known and hence local time. This method is very simple and is independent of any knowledge of latitude and of such instrumental errors as graduation or collimation errors, and is largely free from refraction errors. The disadvantage is that the pair of observations must be taken at a considerable interval of time. If the sun is used a correction must be made for the change in declination in the interval. If a succession of observations is made on different stars, noting the time each reaches the fixed altitude, each observation gives an equation connecting time (hence chronometer error) with altitude, latitude, and declination. The declination is known in each case, and the altitude and latitude are the same in every case, so three such observations give values for time, latitude, and the exact altitude observed. Further observations give a check. Gauss' original solution entailed working out by least squares those corrections to the assumed chronometer error and latitude that make the observations best fit the condition that all the altitudes are equal. MM. Claude and Driencourt have found an extremely simple and accurate graphical solution, which enables the results to be obtained with great speed. The preparation of a programme from the data given in the *Nautical Almanac* entails considerable work, as a suitable number of stars reaching the desired altitude sufficiently close together must be found and their azimuths calculated. This work can, however, be done once for all for any range of latitude, and Messrs. Ball and Knox Shaw have tabulated sidereal times and azimuths of some 17,000 star passages for an elevation of 60°, including only *Nautical Almanac* stars of fourth magnitude and upwards and latitudes from 55° S. to 55° N., based on the star places given in the *Nautical Almanac* for 1918. The tables will in time go out of date owing to change in declination, but they should serve for topographical survey work for some fifteen years. The advantage of the method lies in the fact that an exact knowledge of the altitude is not necessary, hence changes of face are not required with the theodolite, nor determinations of the index error with a sextant. The method can be used with an ordinary theodolite or sextant, but is simplest with the prismatic astrolabe (see article on "Gravity Survey," Vol. III.).

§ (27) Talcott's Method (see article on "Gravity Survey," Vol. III.).—Talcott's method is another excellent method for determining latitude, and can be used with a theodolite provided it be fitted with stops to enable it to be turned exactly 180° in azimuth, and with a micrometer eyepiece.[2] With such an instrument the probable error of a single pair of observations should be about 0·7 second.

§ (28) Determination of Azimuth.—The azimuth of a heavenly body is the angle between the meridian plane of the observer and the vertical plane passing through the body; thus in astronomical work the azimuth is always measured the shortest way from the elevated pole. In triangulation work[3] the azimuth of a terrestrial object is always

[1] See article on "Field Astronomy and Atmospheric Refraction," Vol. III.

[2] See "Description, Adjustments, and Methods of Use, of the 6-Inch Micrometer Theodolite, 1912 Pattern, etc.," *Bulletin 34 of the Canadian Topographic Survey*.

[3] See C. F. Close, *Text-book of Topographical and Geographical Surveying*.

reckoned from the south by the west. If the latitude be known, and also the declination of a heavenly body and its elevation at any moment, its azimuth at that moment can be calculated. To determine the azimuth of a terrestrial object, *i.e.* to determine the direction of the meridian, some reference object (R.O.) is required; this must be a lamp by night, and in any case must be at such a distance that it can be observed to with the telescope at stellar focus. In observing on a star the R.O. is first intersected, and the horizontal circle read, the bottom plate being kept clamped, the telescope is swung on the star, and the horizontal hair set slightly in advance of the star, while the vertical hair is kept on the star by the slow-motion screw. At the moment the star passes the intersection of the cross hairs the slow motion screw is stopped and both vertical and horizontal circles read. It is desirable also to read both the micrometer arm, and the striding, levels. Face is then changed and the star and R.O. again intersected. This completes one set of observations on one star. Observations should always be paired as in other astronomical work. It is also desirable that the star should not be moving rapidly in azimuth, or a small error in the observed altitude will produce a large error in the result. The best time for observing is when the body is in the prime vertical or when a circum-polar star is at elongation.

If local time be known, the azimuth of the star can be calculated if the instant of observation be observed instead of the altitude. The observation in this case is somewhat easier.

In all cases of astronomical observations for azimuth the levelling of the instrument is of great importance, and it is very desirable that the striding level should be used and any dislevelment of the horizontal axis allowed for.

VIII. Levelling and Levelling Instruments

§ (29) Levelling.—Levelling is the art of determining the heights of points on the earth's surface, and the term is generally restricted to the method by which the difference in height between neighbouring points is directly determined by reading with a horizontal line of sight on graduated staves held vertically at the points. If a graduated staff be held vertically at a point A, and read by means of a horizontal line of sight from a point P, the reading gives the difference of height between points A and P. If the staff be now held on any other point B, and again read by a horizontal sight from P, the height of B over A is given by $a - b$, where a and b are the readings on the staves at A and B respectively. The reading a is generally called the "back sight," and the reading b the "fore sight." It is obvious that the points A and B must be comparatively close together, and that the difference in height obtained is with reference to the horizontal plane through P. As the surface of the earth approximates to a sphere rather than to a plane, the reading a is subject to a small correction due to the earth's curvature, and similarly with b. If the distances from P to A and B are equal these corrections will also be equal in magnitude and sign, and consequently the difference $a - b$ will also give the correct difference in height when referred to a spherical level surface. The error introduced by the assumption that the earth is a sphere is quite negligible in levelling operations. If the points whose difference in height is required are so situated that the staves held on them cannot both be read from one position, a series of intermediate points is used, and their heights determined successively.

§ (30) Levelling Instrument.—The levelling instrument consists essentially of some form of sighting apparatus, the line of sight of which can be set in a horizontal plane. The simplest form consists of a water-level. This is illustrated in *Fig.* 13, and consists of

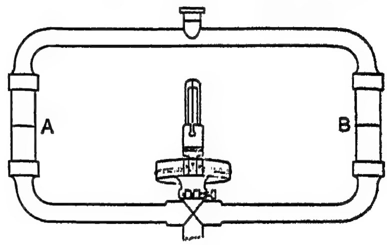


Fig. 13.

two vertical glass tubes A and B connected by horizontal tubes, and mounted on a stand. The tubes are partially filled with coloured water so that the surface of the water is visible in the glass tubes. A line of sight taken across the water surfaces is necessarily level. The instrument is very simple and cannot get out of adjustment. On the other hand, it is impossible to take long or accurate sights.

(i.) *The Ordinary Levelling Instrument.*—This consists of a sighting telescope with a spirit-level attached so that the collimation line can be set horizontal. The important points are that the line of sight of the telescope shall be horizontal when the bubble is in some known position in its tube, and that the telescope can be turned in any direction without altering its height.

The telescope is generally of the astronomical type giving an inverted image, and using an eyepiece of the Ramsden type.

Cross hairs are provided at the focus of the object-glass to define a definite line of sight or collimation line. In order that the line thus determined may not vary as the distance between the object-glass and the cross hairs is altered in the process of focussing, it is necessary that the cross hairs should be accurately on the optical axis of the object glass, and that the mechanical guiding of the moving portion should be true.

Telescopes with an internal focussing lens have been introduced lately. These have the cross hairs rigidly fixed in the tube carrying the object-glass, while the focussing is carried out by means of a negative lens moving between the two. In *Fig.* 14, A is the object

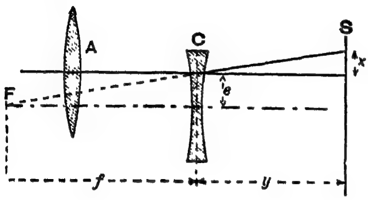


FIG. 14.

glass, C the negative lens, and S the diaphragm carrying the cross hairs. The focal length of C is f, and its axis is supposed parallel to that of A but distant e from it. A ray passing along the axis of A will then be deflected by C and meet the diaphragm at a distance x above the axis. It is seen at once from the figure that $x = ye/f$, where y is the distance between C and S; as in practice f is three or four times y, a given eccentricity in C produces only one-third to one-fourth the change in collimation that would be produced by an equal movement in A or S. Not only is the effect of variable eccentricity reduced by this means, but it is easier mechanically to provide true guides for an internal lens than for the object-glass or diaphragm and eyepiece. A further advantage of this arrangement is that when a glass diaphragm is used, the telescope can be made air-tight, and consequently dust- and damp-proof. An internal positive lens could be used in the same manner, but at the expense of compactness.[1]

Although it is possible to reduce the change in collimation of an instrument to a very small amount by suitable design and good workmanship, it is impossible to eliminate it entirely. For the most accurate work it is possible and necessary so to arrange the observations that errors due to this cause will cancel out.

(ii.) *The Dumpy level* is the type most generally used, and is the simplest type. It consists of a telescope with a spirit-level attached on the top or side, the support of the spirit-level being sometimes cast in one piece with the telescope tube. The whole is mounted on a vertical axis and provided with the usual levelling screws. A clamp and slow motion in azimuth is generally provided for convenience in directing the telescope on the staff. When the instrument is in adjustment the axis of the bubble tube is parallel to that of the telescope and perpendicular to the vertical axis. In use the instrument is set up and focussed and the vertical axis set vertical by means of the levelling screws and bubble in the usual way. If the instrument is in adjustment the collimation line is then horizontal in whatever direction the telescope is turned. The adjustments can be checked and corrected as follows:

(*a*) *Coincidence of Collimation Line and Axis of Telescope.*—Drive in three pegs A, B, and C (*Fig.* 15) so that AB=BC=about 200 feet. Set up the instrument half-way between A and B and ascertain the difference in level between A and B. If the instrument is equidistant from A and B any errors in adjustment will affect the two readings equally, so that the true difference in level will be found. Similarly, find the true difference in level between B and C. Then set up at c as close to A as it is possible to focus the staff, read the staff on pegs A, B, and C.

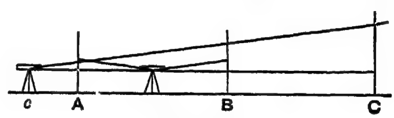


FIG. 15.

From the known correct differences of level compute the correct readings on B and C corresponding to the new reading on A. If the cross hairs are on the telescope axis the error at C will be twice that at B, or more correctly in the ratio cC/cB. If the cross hairs are adjustable any error found can be corrected. If the instrument has been well made and adjusted by the maker no correction should be necessary. As the screws for adjusting the cross hairs (if provided at all) are necessarily small, the less they are used the better, as if they become loose or worn the instrument cannot hold its adjustment.

(*b*) *Parallelism of Bubble Axis and Collimation Line.*—With the instrument at point c (as in above), set the cross hairs at the correct reading on staff at B and then bring the bubble to the centre of its run by means of its adjusting screws. Check the reading on the staff at A, and if this has not altered the adjustment is correct. If a considerable alteration has been made to the bubble screws the reading on A will have been altered, and

[1] See *Zeits. Instrumentenk.*, 1909, p. 320.

the operation must be repeated until the adjustment is correct.

(c) *Bubble Axis Perpendicular to the Vertical Axis.*—Place the telescope parallel to two of the levelling screws of the instrument, and bring the bubble to the centre of its run, then turn the telescope through 180° so as to reverse it end for end. If the bubble is still in the centre of its run the adjustment is correct; if not, correct half the error by the foot screws and half by the screws under the telescope supports, if such be provided. If no arrangement is made for this adjustment the bubble axis must be made perpendicular to the vertical axis, and the collimation line made parallel to the bubble axis by means of test (*b*) and moving the cross hairs. If this introduces any serious error in test (*a*), which is unlikely, the necessary adjustment can only be made by an instrument-maker. With modern methods of manufacture, and with a machined telescope tube, there is no difficulty in making the axis of the telescope perpendicular to the vertical axis with sufficient accuracy for ordinary purposes and without any special adjustment for the purpose. The cross hairs can then be fixed in their correct positions, and the only adjustment that should ever be required to be made by the user is to make the bubble axis perpendicular to the vertical axis.

It should be noted that for ordinary levelling an error of a few seconds of arc in the adjustment is of no importance. For precise levelling it is necessary for a variety of reasons to make the fore and back sights of equal length; and if this be done any adjustment errors cancel out, provided that there is no change of adjustment between the two readings; as in this case it is not necessary to refocus between the readings, the exact collimation is also not necessary. It is seen, therefore, that for precise work also an extremely good adjustment of the instrument is not required.

(iii.) "*Y*" *Levels.*—In this type of level the telescope is provided with two cylindrical collars; the two collars should form parts of the same cylinder, with the same axis as the telescope. The collars rest in Y's, and the bubble tube is attached to the telescope. The telescope can be rotated in the Y's and also reversed end for end after opening the clips which hold it in position. The adjustments are carried out without the use of a staff as follows: (1) The telescope is directed on to some distant object, which is bisected by the horizontal cross hair. The telescope is then rotated in the Y's through 180°. If the object is still bisected, the collimation line is coincident with the axis of rotation of the telescope; if not, half the apparent error is corrected by the diaphragm screws. This brings the collimation line coincident with the axis of rotation. (2) The bubble is then brought to the centre of its run by means of the foot screws. It is then lifted out of the Y's and reversed end for end. If the bubble remains in the centre of its run it is in adjustment; if not, half the error is corrected by the bubble adjusting screws, and half by the foot screws. This brings the bubble axis parallel to the axis of rotation of the telescope in the Y's. (3) The vertical axis is then made vertical in the usual way (see article on "Spirit-levels," § (5)), and if the bubble is not then in the centre of its run it is brought there by means of the screw provided for raising or lowering one of the pillars supporting the Y's.

The above adjustments assume that the axis of the collars coincides with the optical axis of the object-glass. This can be tested by repeating test (1) above on a near object. If the adjustment is correct for a distant object, but not for a near one, the object-glass itself must be moved, and it is not as a rule possible to do this in the field. The adjustment (2) assumes that the collars are of equal diameter. This can be tested by means of a striding level. The striding level is placed on the collars, and the bubble brought to the centre of its run by means of the foot screws. The telescope is then reversed end for end, but the striding level is replaced in its original position. If the bubble is no longer at the centre of its run the collars are of unequal diameters.

The Y level suffers from the disadvantage of a large number of adjusting screws, any one of which is liable to work loose. If the collars become worn or damaged in any way, or if dust or dirt gets between the collars and the Y's, they cannot be used with safety for adjusting purposes, and the instrument is no better than a Dumpy level as regards ease of adjustment, and is more liable to get out of adjustment.

(iv.) *Cooke's Reversible Level.* This is on the same principle as the Y level; but instead of the telescope collars being supported on Y's they are supported in closely fitting rings from which the telescope can be withdrawn longitudinally. The bubble tube is fixed to the rings and not to the telescope direct.

§ (31) PRECISE LEVELS. Levelling instruments for precise work are designed on much the same lines. They are provided with more sensitive bubbles and more powerful telescopes, but the chief difference is that they are provided with a screw for tilting the telescope with the bubble attached, independently of the levelling screws of the instrument, and no attempt is made to set up the vertical axis truly vertical. The bubble is brought to the centre of its run *after* the telescope has been directed on the staff. A mirror or prism arrangement is provided so that the observer

can see that the bubble is set correctly at the moment the observation is taken. This arrangement is very convenient to use, and is to be recommended for all levelling work, except possibly where a large number of staff readings are made from each position of the instrument, as when running cross-sections or when contouring.

(i.) *French Pattern.*[1]—This pattern is of the Y type, and the bubble is fitted as a striding level. Two readings are taken on each staff—(1) with the telescope and bubble in their normal position, and (2) with the telescope rotated through 180° in the Y's and with the bubble reversed end for end.

(ii.) *Coast and Geodetic Survey* (*U.S.A.*) *Pattern.*—This pattern consists of a large Dumpy level with the addition of a tilting screw. The telescope tube is of invar, and the bubble tube is sunk into the telescope tube so as to be as near the telescope axis and as well protected from temperature changes as possible. The bubble is viewed by a system of mirrors and prisms. The telescope is of high power and the bubble very sensitive; in consequence the instrument is heavy. There are only two adjustments, viz. a screw for adjusting the bubble tube parallel to the collimation line, and an adjustment for the cross hairs. It is very suitable for use on level ground where long sights can be taken.[2]

(iii.) *Zeiss Pattern.*—This was designed by Dr. Wild, and has a telescope of special construction. There is no draw tube, and an object-glass is fixed at each end of a rigid tube. The object-glasses are of equal focal length, and each has a set of cross hairs engraved on it. The eyepiece can be attached at either end. The focussing is carried out by an internal lens. The bubble is attached to the telescope tube at the side, and is viewed by the special prism[3] arrangement. The telescope is capable of rotation round its axis through an angle of 180°, and carries the bubble with it. The bubble being ground to a barrel shape can be used in either position. The adjustment is carried out in a special manner. A staff is set up at a convenient distance, and four readings are taken as follows:

(1) Normal working position, bubble to left, prism box up.

(2) Bubble to right, prism box down.

The eyepiece is then placed at the other end of the telescope, which is rotated 180° in azimuth.

(3) Bubble to left, prism box down.

(4) Bubble to right, prism box up.

The bubble is brought to the centre of its run by means of the tilting screw, before each reading is taken. The mean of the four readings is correct, and if the mean is the same as reading (1) the instrument is in adjustment; if not, the telescope is tilted until it reads correct in position (1) and the bubble tube adjusted as necessary. It will be observed that in these four readings two collimation lines are used, the first in positions (1) and (2) and the second in positions (3) and (4). The mean readings (1) and (2), and also the mean of (3) and (4), are free from collimation error. There are also in effect two bubble tubes, the first used in positions (1) and (4) and the second in (2) and (3). The mean of readings (1) and (4) and the mean of (2) and (3) are each free from any error due to the bubble tube not being parallel to the collimation line, hence the mean of all four readings is free from collimation and bubble errors. The method is simple in use, and the instrument holds its adjustment remarkably well owing to the absence of any delicate adjusting screws. It is doubtful, however, whether the complication of the second object-glass is really justified. A simple Dumpy level with a tilting screw, and the bubble mounted and viewed as in this pattern, would appear to be all that is required. The instrument requires adjusting so seldom that the slightly increased labour involved in adjusting by the ordinary method will hardly

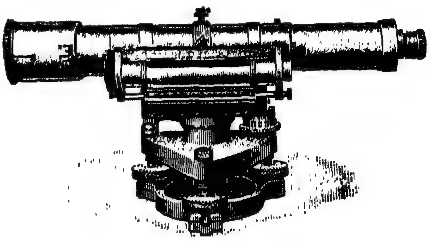


FIG. 16.

be noticed. *Fig.* 16 illustrates such an instrument.

(iv.) *Plate Micrometer.*—It is more accurate to bisect a division on the staff than to estimate the reading of the cross hair when it falls between two divisions. This can be done most readily by mounting a thick plate of glass with parallel sides in front of the object-glass. If the plate be perpendicular to the axis of the telescope it has no effect on the reading, but if it be tilted about a horizontal axis the rays of light are displaced vertically by an amount depending on the thickness and refractive index of the glass and the angle of tilt, hence by tilting the plate the staff can be apparently raised or lowered till the cross hair intersects a mark on the staff. The displacement can be

[1] See Ch. Lallemand, "Nivellement de haute précision," being the third part of "Lever des plans et nivellement" of the *Encyclopédie des Travaux Publics*.

[2] For full description see *C. and G. S. Report*, 1900, p. 521; 1903, p. 200.

[3] See "Spirit-levels," § (10).

read off on a suitably graduated drum connected with the plate. The displacement is read direct, and no calculation involving the distance of the staff is required as is the case when a mark is bisected by tilting the telescope. With a 30-power telescope, and a suitable arrangement of cross hairs, the third place of decimals of a foot can be read accurately up to a range of 100 yards and the fourth place estimated. The form of cross hair shown in *Fig.* 17 is convenient for this purpose, the > being placed over the line on the staff.

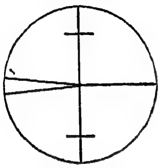


FIG. 17.

§ (32) LEVELLING STAVES.—The usual levelling staff is made of wood, with the graduations painted on. The divisions are either feet and hundredths (sometimes fiftieths), or metres and centimetres or half centimetres. The graduations are usually of the same thickness as the intervals between them, giving a checker pattern. Many different patterns have been devised with a view to ease of reading and avoiding mistakes. The most suitable staff for precise work is probably one with simple line divisions, the divisions just thick enough to be read at the longest range required. This applies especially when a plate micrometer is used, in which case divisions of $\frac{1}{50}$ foot or $\frac{1}{2}$ centimetre are most convenient. As the telescopes generally used are inverting, it is convenient to have the figuring on the staff also inverted. For ordinary work folding or telescopic staves 14 or 15 feet long are convenient, but for precise work they should be in one rigid piece, and 10 feet is long enough. "Target rods" are sometimes used, having a movable target on the face. This is adjusted by the staff holder by signal from the observer at the instrument, so that it is bisected by the cross hair, and the reading is taken by means of a fine scale and vernier on the staff. This method has proved to be neither so convenient nor so accurate as a staff read direct from the instrument, but may be useful where an occasional extra long sight is required.

Wooden staves are liable to change their length with varying humidity; a variation of 3 to 5 parts in 10,000 may be expected from this cause. This is serious in precise work in hilly country. If such staves be used their length must be checked at regular intervals either by a standard bar or otherwise. A method due to Colonel Goulier, of the Commission du nivellement général de la France, is as follows: Two bars of iron and brass respectively are run up the centre of the staff, being secured at the foot of the staff and free to slide elsewhere. They carry at their upper ends marks which are read against a scale fastened to the wood of the staff. The combination of the two rods forms a metallic thermometer, so that from the readings their true lengths and also that of the staff can be found. Since the introduction of invar one bar of this material has been substituted for the two of iron and brass. Staves graduated direct on invar have been used by the Ordnance Survey; there was considerable difficulty in obtaining these, but the Cambridge Scientific Instrument Co. finally provided some of such accuracy that the staff readings could be used direct without correction for temperature or graduation errors.

If the staff be not held vertical when the reading is taken the reading will always be too high. A common method of dealing with this is to swing the staff and estimate the lowest reading of the cross hair; this, however, decreases the precision of the reading. It is equally quick and much more accurate to hold the staff vertical with the assistance of a small circular spirit-level. In this case some means of holding the staff steady must be employed; that illustrated in *Fig.* 18 is convenient.

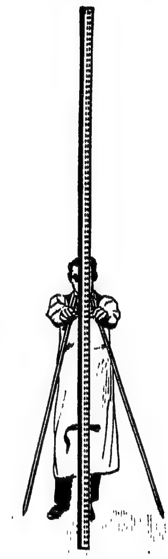


FIG. 18.

§ (33) STAFF SUPPORTS.—The stability of the support on which the staff rests is of great importance; it is obvious that if this changes its height between the forward reading to the staff from one position of the instrument and the back reading from the next position, an error will be introduced. For ordinary work it is convenient to use a small metal tripod pressed into the ground; for precise work a peg (either wood or metal) driven in is generally used, but the actual support used must depend on the nature of the ground. The error caused by the instability of the staff supports is one of the most important in long lines of precise levelling.

§ (34) BENCH MARKS.—The ordinary bench mark of the Ordnance Survey consists of a horizontal line cut in some vertical face, with the Government "broad arrow" cut below it, thus ⊼. In the new precise levelling the marks are of two kinds. The "fundamental" marks, which average 30 miles apart, are illustrated in *Fig.* 19; they are confined to sites where the reference points can be either

fixed in the living rock or in concrete founded on the rock. There are three actual reference points—one a gun-metal bolt for general use,

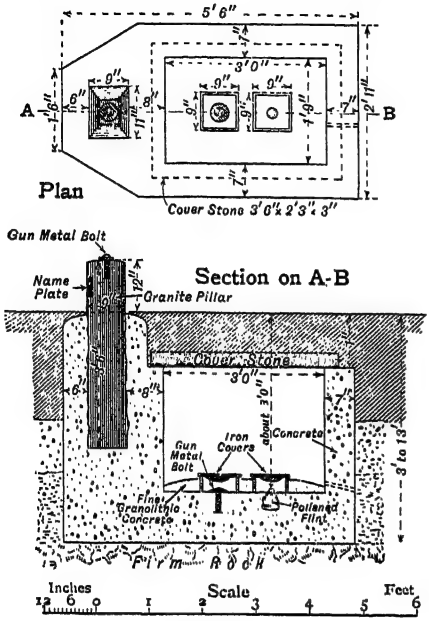


FIG. 19.

and the other two buried for the future use of the survey. The secondary marks, generally about a mile apart, are gun-metal plates as shown in *Fig.* 20, cemented into a vertical

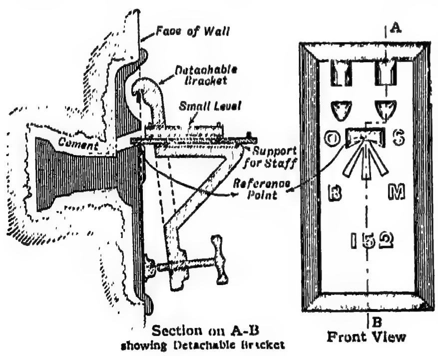


FIG. 20.

face; a detachable bracket serves to support the staff.

§ (35) ERRORS IN LEVELLING.—Some of the causes of errors have already been touched on. Errors due to lack of adjustment of the instrument are eliminated if the instrument is equidistant from the two staves, provided the adjustment does not alter between the two readings. Such change may be due to temperature effects or change of focus. The former acts slowly, provided the instrument is shielded from the direct rays of the sun; the latter does not come in if the instrument is equidistant from the staves. If there is any tendency for the staff supports to move always in the same direction a systematic error will be introduced, which in long lines may easily become the predominant error. The only way to eliminate this is to relevel the line in the reverse direction, keeping all other conditions as unaltered as possible; this is invariably done in all precise work, and a systematic difference between the two levellings has almost always been found. Another cause of error is due to atmospheric refraction; the statement in many text-books that this error cancels out when the instrument is equidistant from the two staves is only true on the assumption (generally incorrect) that the atmospheric conditions along the paths of the two rays are the same. See article on "Trigonometrical Heights and Atmospheric Refraction," Vol. III.

§ (36) DEFINITION OF HEIGHT.—The obvious definition of the height of a point is the distance of the point above the reference surface measured along the vertical through the point. This is called the "orthometric" height. As the surface of the earth approximates to a spheroid rather than to a sphere, this definition leads to the result that two points at the same height are not necessarily at the same gravitational potential, and water might flow from one point to the other, or even from a lower point to a higher. An alternative definition is based on the assumption that all points which are at the same "height" are on the same equipotential surface, and measuring the actual height of each such surface at some fixed point. For this purpose the earth is generally assumed to be a spheroid, and the fixed point is taken in latitude 45°. This is called the "dynamic" height. In practice the distinction only becomes of importance when large areas are covered with a network of precise levelling, and especially when the area covers a tract of hilly country extending north and south. The difference of height as determined by spirit-levelling does not conform to either of the above definitions. The height is first measured on a vertical staff (orthometric) and then carried horizontally to a second staff (dynamic), and so on. The result is that a circuit of levelling carried out without any error, and closing on the starting-point, will, in general, show a difference in height between the starting-point and the (same) finishing-point. The magnitude of this apparent error can be calculated on the assumption that the vertical as indicated

by the level at each setting up of the instrument is a normal to the spheroid of reference. In practice it is the normal to the geoid, and the true correction to be applied to the levelled heights to give either orthometric or dynamic heights cannot be calculated unless the deviation of the plumb-line, or the true variation in the force of gravity, be known for each station occupied. F. R. Helmert[1] has calculated the theoretical closing error of a circuit in the Tirol (*a*) assuming the usual formula, (*b*) based on the actual values of gravity measured by R. V. Steineck; the results were 7 mm. and 24 mm. respectively.[2]

IX. Contouring

§ (37) Contouring.—A contour is a line of equal height above sea-level, and contours drawn on a map at equal vertical intervals form the best method of showing the vertical relief of the ground. The vertical interval depends on the scale of the map and the steepness of the ground. On small-scale maps the contours are necessarily generalised and become "form lines." Accurate contouring must be controlled by spirit-levelling. In large-scale work the contours are established on the ground by means of a spirit- or water-level, and surveyed. The survey may be done either simultaneously with the marking of the contour, or later, and may be done by chain survey, stadia traverse, or plane-tabling, or by a combination of the above.

In contouring with a plane-table on small scales the Indian pattern clinometer is a useful instrument. It is 9 in. long, and has two vanes which are upright when in use. The rear vane has a sight-hole, and the front vane a vertical slit with a scale of degrees on one side of the slit and of natural tangents on the other. The sight-hole is brought level with the zero of the scales by means of a screw, a level on the horizontal arm indicating the position. The adjustment must be checked by comparison with the readings of a theodolite or by means of reciprocal observations. It is used standing on the plane-table, and by its means the height of the table can be found by measuring the vertical angle to any point whose height is known. The slope of the ground and hence the spacing of the contours on the map can also be found. Sights can be taken up to a distance of about 3 miles. By fixing the height of ruling points, and the slope down spurs and valleys, etc., the contours can be drawn in with sufficient accuracy for small-scale work.

[1] *Die Schwerkraft in Hochgebirge*, p. 19.
[2] Ch. Lallemand, *Nivellement de haute précision*; *Coast and Geodetic Survey Reports*, 1900, App. 6, 1903, App. 3; *Account of the Operations of the G. T. Survey of India*, xix.; E. O. Henrici, *Proc. Inst. Civil Engineers*, ccix. Pt. 1.

X. Base-line Measurement

§ (38) General.—Before a triangulation can be computed, it is necessary to know the length of one of the sides of one of the triangles, and this must be measured directly. This measured side is called the base, and thus forms the starting-point of any system of triangulation required to determine the horizontal co-ordinates of its points. These co-ordinates are measured on the spheroid of reference, consequently the length of the base (as required for purposes of calculation) is the distance, measured along the spheroid, between the points on the spheroid vertically below the ends of the actual base. Thus if A, B (*Fig.* 21) are the terminal points of the base, ANB being the ground surface, the distance required is A′B′ (measured along the curve) where A′B′ lie on the spheroid of reference and AA′, BB′ are the verticals through A and B.

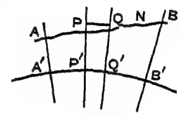


Fig. 21.

The actual line followed by the base may be either in the plane through AA′B or that through ABB′, or it may be such that if P be any point in the base and PP′ the vertical through it, ABPP′ all lie in one plane. These cases represent respectively the result of the line being laid out by a theodolite at A, or at B, or by a transit theodolite at each consecutive point P. The three lines are not strictly the same when the spheroidal shape of the earth is taken into account, but the differences are negligible for base-line measurements, which only extend for a few miles, hence any of these three methods may be used for laying down the base on the ground. The base is necessarily measured in small sections such as PQ, representing the length of a single bar or tape. As these are short compared with the radius of curvature of the spheroid, the straight line distance PQ can be taken as equal to the arc, provided the length of PQ is horizontal. If PQ be not horizontal it must be multiplied by cos i, where i is the inclination of PQ to the horizontal. Unless the base is being measured at sea-level, a further reduction must be made. If R is the radius of curvature of the spheroid at P′ and h the height of P above P′, we get

$$\mathrm{P'Q'} = \frac{\mathrm{R}}{\mathrm{R}+h} \cdot \mathrm{PQ} \cos i,$$

whence L, the length of the base, is given by

$$\mathrm{L} = \Sigma(\mathrm{P'Q'}) = \Sigma(\mathrm{PQ} \cos i) - \Sigma\left(\mathrm{PQ} \cdot \frac{h}{\mathrm{R}}\right),$$

or if i is small, as it is when rods are used,

$$\mathrm{L} = \Sigma(\mathrm{PQ}) - \tfrac{1}{2}\Sigma(\mathrm{PQ} \cdot i^2) - \Sigma\left(\mathrm{PQ} \cdot \frac{h}{\mathrm{R}}\right).$$

These two corrections are known as the reduction to horizontal and reduction to sea-level respectively. It is not necessary in practice to correct each length PQ separately for reduction to sea-level, and the whole correction can be put in the form $L(H_m/R_0)$ where H_m is the mean height of the base above sea-level, and R_0 the mean radius of curvature of the spheroid from A′ to B′.

§ (39) Historical.—The earlier bases were frequently measured by means of deal rods shod with metal at their ends. Such rods do not vary greatly in length with change of temperature, but alter their length considerably with varying moisture. This variation can be largely reduced by boiling the rods in paraffin wax. Wooden rods were used in the classic measurements undertaken by the French in Lapland and Peru in 1735–45. They have been used successfully in Germany,[1] and more recently in Egypt,[2] where a probable error of about five parts in a million was obtained by this method.

The earliest base in this country (Hounslow Heath,[3] 1784) was measured with deal rods with bell-metal ends, and with glass rods. In the case of the deal rods it was noticed during the work that their length was much affected by changes in the humidity of the atmosphere. No attempt appears to have been made to avoid this by varnishing or otherwise treating the rods. The temperature of the glass rods was found by two thermometers in contact with each rod, and the expansion due to temperature was allowed for. The discrepancy between the two measurements was 1·8 ft. in 27,400 ft., or 66μ (μ in this section represents one-millionth of the total length). In 1791 this base was remeasured with a steel chain, made by Ramsden, of 100 ft. length (40 links of 2·5 ft. each); the discrepancy of the result compared with the glass-rod measurement was 0·23 ft., or 8μ. The Ramsden chain was used for four other bases in England, and is now deposited in the Science Museum at South Kensington.

During the nineteenth century accurate bases were almost invariably measured with metal bars, a probable error of 1 to 2μ being obtained. In 1885 E. Jäderin [4] suggested the use of metal tapes or wires suspended free in catenary for base measurements, and the introduction of "Invar," due to the researches of Ch. Ed. Guillaume, of the International Bureau of Weights and Measures, has enabled Jäderin's apparatus to be so improved that it now forms a method of base-line measurement which is not only fully as accurate but much quicker and cheaper than the older methods.

The measurement of the Salisbury Plain base (length 6·9 miles) in 1849 with Colby's apparatus took a large party nearly six months in the field, while the Lossiemouth base of 4·5 miles was measured in 1909, using invar tapes, by a much smaller party in about a month, and with a higher degree of accuracy.

§ (40) Base Measurement in the Nineteenth Century.—The great trouble in all base line measurements has been the change in length of the apparatus due to humidity (in the case of wood bars) or temperature. A metal rod needs time to take up the temperature of its surroundings, and in the field the conditions are such that the true mean temperature of the rods cannot be obtained satisfactorily by the use of thermometers in contact with them. To overcome this difficulty Borda [5] introduced the principle of the bimetallic thermometer. His apparatus consisted of two strips of metal in contact. The lower strip of platinum (either 2 toises or 4 metres long) rested on a stout beam of wood; lying immediately on it, and fastened to it at one end, was a strip of copper about 6 inches shorter. A scale on the free end of the copper, read by a vernier on the platinum, indicated the relative expansions of the two bars, whence the temperature, and therefore the true length of the platinum strip, could be inferred.

Another apparatus designed to the same end was Colby's [6] compensated bar apparatus, which consisted of two 10-ft. bars, one of iron and one of brass, firmly connected at their centres. At either end was a metal tongue about 6 inches long pivoted to both bars so that, while free from shake, they did not impede the free expansion of the bars. At the outer end of each tongue was a small dot, and the distance between the dots determines the length. If aa' and bb' (*Fig.* 22) are the bars,

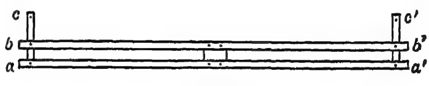


Fig. 22.

having a temperature coefficient of α and β respectively, and if abc, $a'b'c'$ are the tongues, the lengths ab, bc, etc., are so chosen that $ac/bc = \alpha/\beta$; if, therefore, owing to a change of temperature, a moves a distance ad, b will move βd, and c will not move. If the expansion can be considered as proportional to the change of temperature, and if both bars are at the same mean temperature, the distance

[1] Reinhertz, *Zeit. Vermess.*, 1806, Pt. 7.
[2] Lyons, *Cadastral Survey of Egypt*, Cairo, 1908.
[3] *Roy. Soc. Phil. Trans.*, 1785; also *Account of the Principal Triangulation of the Ordnance Survey*, 1858, p. 206.
[4] Jäderin, *Geodätische Längenmessung mit Stahlbändern und Metalldrähten*, Stockholm, 1885. An English translation appears in the *C. and G. S. Report for* 1893, Pt. II. App. 5.

[5] Delambre, *Base du système métrique, etc.*, Paris, 1806–10, ii. 1 and iii. 311.
[6] H. Yolland, *Measurement of the Lough Foyle Base*, etc., London, 1847.

cc' will remain constant. This apparatus was used for the two chief bases of the Ordnance Survey of the United Kingdom, for ten bases in India, and also in South Africa. It has not been used since 1870, as it was found that the temperatures of the two bars did not remain equal, and it was found necessary in India to use thermometers, and not to rely entirely on the automatic compensation.[1] The probable error of measurement with this apparatus was about $1{\cdot}5\mu$.

In most of the earlier bases the rods were used as end measures, and were laid touching end to end. Borda introduced a graduated slider moving in a groove at the free and uncovered end of the platinum strip, for the purpose of measuring the small gap left between successive bars. This was improved later by the addition of a spring and contact screw. Struve,[2] for the Russian bases, used a wrought-iron bar (whose temperature was determined by two thermometers) fitted with a contact lever (*Fig.* 23) for measuring the small interval between the bars, the lever being held in contact by a spring. Bessel[3] used bimetallic bars of iron and zinc. The upper (zinc) terminated in horizontal knife-edges, while at the free end a small piece with two vertical knife-edges was attached to the iron bar (*Fig.* 24). The distance between the knife-edges was measured by a glass wedge, and gave the temperature correction, and the distance between successive bars was measured in the same way. The bases measured with this apparatus had an error of about 2μ.

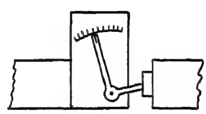


FIG. 23.

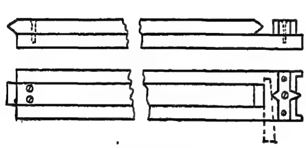


FIG. 24.

Bache in 1845 designed an apparatus which was used by the Coast Survey of the United States, and which combines several of the above-mentioned arrangements. *aa'*, *bb'* of *Fig.* 25 are iron and brass rods respectively, forming a Borda bimetallic thermometer. They are separated by rollers so that they can expand freely. The cross-sections of the bars are so arranged that, while they have equal absorbing surfaces, their masses are inversely proportional to their specific heats, allowance being made for their difference of conducting power. Lever C is a compensation lever so arranged (as in Colby's apparatus) that the position of the upper knife-edge remains constant irrespective of the expansion of the bars. A collar on a sliding rod *d* is pressed back against the knife-edge by a spring, and the sliding rod terminates in an agate plate for contact with the next bar. A scale and vernier attached to the free ends of *a* and *b* enable the relative expansion of the bars to be measured if desired. At the other end, where the bars are fixed together, is another sliding rod *e* terminating in a blunt knife-edge for making contact with *d*; the inner end of *e* abuts on a contact lever *f* pivoted below, which in its turn comes in contact with the short tail of the lever *g*, mounted on trunnions, but not balanced, and carrying a spirit-level. For a certain position of the rod *e* the bubble becomes central, and this position determines the standard length of the bar. This arrangement is very delicate, and ensures that the pressure of contact is always the same.

All the bars so far considered (except Colby's) are end measures. Colby set the bars so that the dots of two successive bars were exactly 6 inches apart, by means of two microscopes. The two microscopes were connected by two bars of brass and iron respectively in such a manner that the focal points of the microscopes were exactly 6 inches apart, the temperature compensation being on the same principle as that of the bars. A third microscope passed through the centre of the bars, and was constructed to focus a point at ground-level. The microscopes were provided with levelling screws, and screws for giving motion in longitudinal and transverse directions. The whole compensated microscope set rested on a bracket attached to one end of the case enclosing the bars. The measurement of the 6-inch intervals was effected by bringing one of the outer microscopes to intersect the dot on the bar last placed, and then moving the next bar up by its slow-motion screws so that its dot was

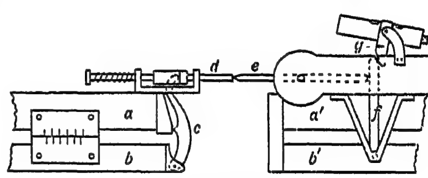


FIG. 25.

[1] *G. T. Survey of India*, Dehra Dun, 1870, i.

[2] F. G. W. Struve, *Arc du Méridien de* 25° 20′ *entre le Danube et la mer glaciale mesuré depuis* 1816 *jusqu'en 1855, sous la direction de Tenner, Selander, Hansteen, et F. G. W. Struve*, St. Petersburg, 1857–60.

[3] F. W. Bessel and J. J. Baeyer, *Gradmessung in Ostpreussen, etc.*, Berlin, 1838.

bisected by the other microscope. The end of each section of six bar lengths was transferred to the ground by the centre microscope, a suitable reference point being adjustably mounted on a heavy cast-iron plate resting on the ground. The Salisbury Plain base, which was measured in 1794 with the Ramsden chain, was remeasured in 1849 with the Colby apparatus, and a discrepancy of 1·03 ft., or 28μ, was found between the two measurements. Some of the original Colby apparatus is now in the Science Museum at South Kensington.

Boscovitch in Italy in 1751 used bars with the length defined by line marks on the bars, measuring the distances between successive bars by dividers and scale, and other early work was done with marks on the sides of the bars.

All the above-described methods used a succession of bars, either actually or nearly touching. J. Porro and others used a single bar to measure the distance between successive microscopes mounted on tripods, and carefully aligned and levelled. Porro used a bimetallic bar for temperature correction. His apparatus was improved by Ibañez [1] and used in Spain and Switzerland.

In 1871 R. S. Woodward,[2] of the Coast and Geodetic Survey, overcame the temperature difficulty by using a single iron bar carried in a V-shaped trough filled with melting ice. An error of $0{\cdot}25\mu$ is claimed for this apparatus, and it is probably the most accurate method that has been adopted. It is slow and expensive to use, and its greatest utility is perhaps for laying down a field standard for standardising wires or tapes.

Any method, such as those described above, which involves the use of metal measuring bars in the field entails the provision of heavy trestles for their support. These trestles must be carefully aligned and levelled, and in order to do this the ground must be fairly level and smooth, and in many cases much clearing and preparation of the ground are required. Such methods are therefore necessarily slow and expensive.

For further particulars of the older methods see A. Westphal, *Zeits. Instrumentenk.*, 1885, v. 257, 333, 373, 420, and 1888, viii. 189, 225, and 337. For fuller details reference must be made to the original accounts mentioned, or to the official reports of the surveys concerned.

§ (41) Modern Base Measurements. (i.) *In America.*—The American "Duplex" bar apparatus has been used with success in modern work. It consists [3] of two measuring tubes 5 metres long, each containing two hollow cylindrical bars, one of steel and one of brass. The outer tube is mounted on two wooden tripods, with fine adjustments for levelling and aligning. The measuring bars project at both ends from the outer tube, and are brought into end contact, steel to steel and brass to brass, by means of screws which move them bodily in the outer tube. The contacts are agate knife-edges. The process results in two simultaneous, but independent, measurements, one with the steel bar and one with the brass bar. If, owing to temperature, one bar gain too much on the other, it can be set back a known amount by means of a vernier on the steel bar read against a scale on the brass. Thermometer readings give the temperatures of the bars. The difference between the total lengths of the base as obtained from the brass and steel, uncorrected for temperature, gives the mean temperature of the bars, and thus the temperature correction to be applied. The thermometers give the corrections to be applied to each individual bar length, which provides a check. The bars are actually tubes of comparatively thin metal, the thicknesses being calculated to compensate for relative specific heats and conductivities; they are both nickel-plated, and enclosed in a double tubular covering; the result is that they maintain very closely the same temperature.

(ii.) *Tapes.*—If wires or tapes be used for the actual field measurements, they can be much longer than the bars, consequently there are fewer points requiring accurate alignment, and fewer measurements or settings at the end of each unit length; this tends to both speed and accuracy. At the measurement of the Holton Heath base [4] a comparison of various methods was made and the practical use of tapes tested in the field. Steel tapes graduated every 20 metres over a length of 100 m. were employed, the total length being about 101 m.; the cross-section was 6·34 by 0·47 mm., and the weight 22·3 grams per metre. The tape was strained to a constant tension of 25 lbs. 9 oz. by means of a spring balance. The tensioning apparatus consisted of a lever of steel tubing, with a wooden extension as handle, hinged by a universal joint to a platform on which the operator stood. The tube was threaded, and a wheel nut carried a gimbal-jointed support, to which was attached the spring balance. The balance was used at one end only. The tape was supported every 20 m. on wire nails driven into posts. At the end of every 100 m. length a properly ranged marking-post was cut off, or driven down, to the proper height, and capped by a plate of zinc. The supporting

[1] C. Ibañez et Saavedra, *Expériences faites avec l'appareil à mesurer les bases appartenant à la commission de la carte d'Espagne* (Traduction par A. Laussedat, Paris, 1860).

[2] R. S. Woodward, "On the Measurement of the Holton Heath Base, etc," *Report of the Coast and Geodetic Survey for 1892*, Pt. 2, App. 8.

[3] *Coast and Geodetic Survey Report*, 1897, App. 2.

[4] R. S. Woodward, "On the Measurement of the Holton Heath Base, etc.," *Report of the Coast and Geodetic Survey for 1892*, Pt. 2, App. 8.

nails were carefully adjusted as to height by a telescope, so that they lay in a line joining the two marking-posts. The tape being placed in position, the rear mark was brought to coincide with a mark on the zinc top of the marking-post, and the strain taken at the front end by the spring balance; the tape was then lifted clear of the nails, and dropped on to them so as to avoid friction and any lack of alignment in the tape itself. When all was correct, a scratch corresponding with the forward end of the tape was made on the zinc of the forward marking-post with a fine bradawl, the thermometers were read, and the tape moved forward to the next section. If temperature effects, or inaccurate spacing of the marking-posts, caused the tape ends to get too near the edges of the zinc plates, the tape was set forward or back an even number of centimetres by means of a scale. The tapes were standardised in the field on a special 100 m. comparator by means of the 5 m. iced bar. A special section of the base 1 km. long was measured by the iced bar, by 5 m. contact rods, and by the 100 m. tape. The tape measurements were taken both by day and by night. Two or more measurements went to a set; the sets taken at night varied $\pm$ 3 mm. from the mean, those by day up to 10 mm. The mean of the night sets (21 measurements) differed from the iced bar measurement by only 0·2 mm. The greater accuracy at night is due to the better temperature conditions. The rate of working with the iced bar was 100 m. an hour, and the probable error of one such measurement of 100 m. was about $\pm$ 0·034 mm. The probable error of the 1 km. section (mean of 4 measurements with the iced bar) was $\pm$0·26 mm. The rate of working with the tapes at night was about 2 km. per hour. The probable errors given are those of measurement only, and do not include the error of the iced bar itself. The final results for the kilometer section were:

Mean of 6 measurements with steel rods	1000 m. – 4·9 mm.
Mean of 21 measurements with steel tapes	1000 m. – 3·6 mm.
Mean of 4 measurements with iced bar	1000 m. – 3·4 mm.

These results showed the possibility of using steel tapes for work of first-class accuracy.

In 1900 the Coast and Geodetic Survey [1] considered that, in view of the speed with which base measurements could be made with tapes, it would be advantageous to measure frequent check bases. An accuracy of 2 μ was laid down, and no time was to be spent in trying for more accurate results than this. Fifty and 100 m. tapes were used, supported every 25 m.—2 × 4 in. stakes for intermediate supports, and 4 × 4 in. stakes for the marking-posts, topped with copper instead of zinc. A kilometer of each base was measured with the duplex bar apparatus, and tapes and bars were standardised against the iced bar at the beginning and end of the season. The methods for using the tapes were generally the same as those employed at the Holton Heath base. In the result, 9 bases with a total length of 69 km. were measured by one field party between July 16 and January 18; including the standardisation work, the worst probable error obtained was 1·4 μ and the average 0·8 μ. Later on invar tapes were used,[2] which enabled work to be done in the daytime with less danger of temperature errors and greater ease and accuracy, as compared with steel tapes by night. In 1907 a party of from seven to thirteen men prepared and measured 3 bases, and measured 3 more (prepared by another party), in all 6 bases with a total length of 100 km., in four months. The worst probable error was 0·46 μ and the average 0·36 μ. Both steel and invar tapes were used, the latter proving 7 per cent faster and 70 per cent more accurate.

(iii.) *Tapes or Wires in Catenary.*—When tapes are used as above described, with the tapes supported at short intervals, and always supported in the same way both when being standardised and when in use, and when only small inclinations are used, no allowance need be made for sag, etc. The method, however, while much quicker than the use of bars, and not necessitating quite such a smooth stretch of ground, still requires a fairly smooth and level site for the base. The latest base apparatus, a development of Jäderin's,[3] uses the tapes hanging free in catenary throughout their length. By this method fewer supports are necessary, and bases can be accurately measured over more uneven ground and on steeper slopes.

The method and apparatus used have been described in detail by Benoit and Guillaume.[4] The tapes are not graduated throughout their length, but carry small scales at either end, reading + or − from the zero graduation. The scales are a few inches long and are graduated in millimetres, or some similar divisions such as $\frac{1}{25}$ in. The nominal length of the tape is the straight line distance between the zero graduations when they are at the same level, with the tape hanging free at standard tension. There is some difference of opinion

[1] *Coast and Geodetic Report*, 1901, App. 3, p. 220.

[2] *Coast and Geodetic Survey Report*, 1907, App. 4, p. 107.

[3] Jäderin, *Geodätische Längenmessung mit Stahlbändern und Metalldrähten*, Stockholm, 1885.

[4] Benoit and Guillaume, *La mesure rapide des bases géodésiques* (1st ed., Paris, 1905; 5th ed., 1917). See also *Ordnance Survey Professional Papers*, New Series, No. 1, 1912, and *Measurement of an Arc of Meridian in Uganda* (Colonial Survey Committee Report), 1912.

as to the relative merits of tapes or wires. If tapes are used, the scales can be marked direct on the tapes themselves, and it is easy to ensure that there is no twist in the tape. With wires the scale has to be attached to the wire, and a twist is not so easily detected. On the other hand, wires are probably less affected by wind. The tension is applied to the tape by means of a cord or wire, passing over a pulley, with a weight attached. The measurement is carried out as follows: A number of trestles are aligned along the base and spaced at intervals of one tape-length. The trestles have heads somewhat as shown in *Fig.* 26, the heads being movable horizontally on the trestles for accurate alignment. The tape is then placed alongside the trestles, just not touching the tops, and strained by means of the straining trestles to the standard tension. Simultaneous readings are taken by observers at the two ends, the scale on the tape being read against the mark on the trestle. Eight or ten such readings are taken, the tape being moved slightly in a longitudinal direction between each reading. The tape is then moved to the next span and the process repeated. After every six spans or so the observers change ends, in order to eliminate personal errors. Each section of a convenient number of spans is measured two or more times, the direction of measurement being changed between each. It is preferable to use more than one tape. At the end of each day's work the point on the ground directly under the terminal trestle head is marked on the ground, and checked before recommencing work to ensure that the trestle has not shifted. The difference in height between the trestle heads must be ascertained by levelling, or by observing the slope between them. This is required both for the ordinary reduction to the horizontal and to allow for the change in shape of the catenary when the ends of the tape are no longer at the same level. The temperature of the tapes must also be noted; but as the coefficient of expansion of invar is so small (about 5×10^{-7} per 1° C., or less), no great accuracy is necessary. The temperature can be observed either by means of a swing thermometer or by thermometers attached to the tape beyond the terminal marks; the former is probably the preferable method, at any rate in temperate climates.

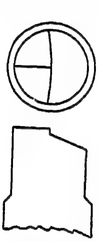


FIG. 26.

The apparatus as originally designed had weights and pulleys at both ends for applying the tension; but if the tape is on any appreciable slope the tension for equilibrium is not the same at the two ends, and the tape will run downhill, or, if it does not, the excess tension at the lower end must be absorbed by friction, and in such a case an uncertainty as to the true tension at once arises. A preferable method is to attach one end direct to the straining trestle by means of a screw, and to move the tape by its means between the readings, half the readings being taken after a screwing-up motion and half after a screwing-out motion; if a fine wire or steel tape be used for connecting the weight, there is practically nothing lost by friction (which is not necessarily the case when a cord is used), and the mean tension applied at the weighted end is the standard tension as determined by the weight.

No elaborate apparatus is required. The straining trestles for the tapes consist of one long leg carrying the pulley, which is placed in the alignment of the base, and two shorter legs supporting it, one on each side. If the long leg be continued beyond the point of attachment of the supporting legs, so as to form a handle, it can be placed in its correct position without much difficulty. In the simplest form the pulley is mounted in a slot in the leg, but some users prefer having it mounted in a frame, the frame being pivoted so that the pulley automatically places itself in the plane of the tape, being capable of movement up or down, and sideways, relative to the trestle, by means of screws. In any case it is essential that the pulley should be true on its axis, and as frictionless as possible.

The trestles for marking the ends of the tape lengths are easily aligned by a theodolite. The difference of height between successive trestles is best determined by an ordinary telescopic level, using a special short levelling staff resting on the trestle heads. It is sufficient in the field to read the scales on the tapes with the aid of a reading-glass; microscopes are not necessary.

Great care must be taken in handling the tapes or wires. It must be remembered that the whole accuracy depends on the constancy of their lengths; they must therefore on no account be strained beyond their elastic limit, or bent or kinked. To ensure this, great care is necessary in winding and unwinding the tapes on the drums used for storage or transport, and they should never be dragged along the ground. The following points must be kept in mind in the design and use of the drums: (1) The diameter of the drum must be such that no sharp bend is given to the tape or wire; a diameter of about ½ metre is suitable. (2) In the case of a wire, which has a stiff scale attached, the end must be attached to a projecting hook on the drum of such length that the scale forms a tangent to the drum, touching the drum well beyond the

scale, so that the wire is not bent at the point of its attachment to the scale. (3) Wires should be wound helically on the drum without overriding. (4) Tapes should be wound on a narrow drum with cheeks, the distance between the cheeks being only slightly greater than the width of the tape to avoid any possibility of the tape catching in itself when being unwound. (5) The tape or wire must not be wound so tightly that any expansion of the drum with temperature can possibly strain the wire beyond its elastic limit. There is no difficulty in the case of a tape, but if a wire is not rolled tightly the turns may override, and it is safer to insert a spring between one end of the wire and its hook on the drum, and thus ensure a sufficient but not excessive tension. Provision must be made for mounting the drum, either on the straining trestle or in its carrying box, so that it can be easily turned for winding and unwinding.

It is desirable to keep one or more tapes as standards, and to use one or more working tapes for the actual base measurement. A pair of marks similar to those on the trestles should be set up one tape-length apart on firm-fixed pedestals somewhere near the base, for use as a comparator for comparing the working tapes with the field standards, and such comparison should always be made before and after the base measurement, even if intermediate comparisons are not made. By this means any variation in the length of the working tapes can be detected. These latter cannot be so effectively guarded against accidental damage as the standards.

§ (42) Theory of Tapes in Catenary.—If a heavy, uniform, and inextensible string is stretched between two points the equations of the curve assumed are

$$x = cu,$$

$$y = c\cosh u = c\left(1 + \frac{u^2}{2!} + \frac{u^4}{4!} + \dots\right),$$

$$s = c\sinh u = c\left(u + \frac{u^3}{3!} + \frac{u^5}{5!} + \dots\right),$$

where s is the length measured along the curve from the lowest point, c is the length of string whose weight is equal to the tension at the lowest point, and u is an independent variable; also the tension at any point on the curve is equal to the weight of a piece of string equal in length to the y of that point. Hence if w is the weight per unit length, T_0 the tension at the lowest point, T_1 and T_2 the tensions at points (x_1y_1), (x_2y_2) respectively, then $c = T_0/w$, $y_1 = T_1/w$, $y_2 = T_2/w$.

If l is the length of the string $(= s_2 - s_1)$,
X the horizontal distance $(= x_2 - x_1)$,
h the difference in height between the ends $(= y_2 - y_1)$,

the problem is to find X in terms of X_0 (the value of X when $h = 0$) and h.

$$X = x_2 - x_1 = c(u_2 - u_1),$$

$$l = s_2 - s_1 = c(\sinh u_2 - \sinh u_1)$$

$$= 2c\cosh\frac{u_2 + u_1}{2}\sinh\frac{u_2 - u_1}{2}.$$

Similarly

$$h = 2c\sinh\frac{u_2 + u_1}{2}\sinh\frac{u_2 - u_1}{2}$$

and $$\frac{l^2 - h^2}{4c^2} = \sinh^2\frac{u_2 - u_1}{2} = \sinh^2\frac{X}{2c},$$

call $$\frac{l^2 - h^2}{4c^2} = z.$$

Then

$$X = 2c\sinh^{-1}\sqrt{z},$$

$$= 2c\sqrt{z}\left(1 - \tfrac{1}{6}z + \tfrac{3}{40}z^2 - \tfrac{5}{112}z^3 + \dots\right)$$

$$= \sqrt{l^2 - h^2}\left(1 - \tfrac{1}{6}z + \tfrac{3}{40}z^2 - \tfrac{5}{112}z^3 + \dots\right).$$

Also $$y^2 - s^2 = c^2(\cosh^2 u - \sinh^2 u) = c^2,$$

hence $$y_2^2 - y_1^2 = s_2^2 - s_1^2,$$

and $$l(s_2 + s_1) = s_2^2 - s_1^2 = y_2^2 - y_1^2 = h(y_2 + y_1),$$

$$l(s_2 - s_1) = l^2,$$

hence $$2ls_1 = h(y_2 + y_1) - l^2,$$

$$2ls_2 = h(y_2 + y_1) + l^2,$$

and $$2l^2(s_2^2 + s_1^2) = h^2(y_2 + y_1)^2 + l^4;$$

but $$2c^2 = (y_2^2 + y_1^2) - (s_2^2 + s_1^2),$$

hence $$4l^2c^2 = 2l^2(y_2^2 + y_1^2) - h^2(y_2 + y_1)^2 - l^4$$

$$= l^2(y_2 + y_1)^2 + l^2h^2 - h^2(y_2 + y_1)^2 - l^4$$

$$= (l^2 - h^2)\{(y_2 + y_1)^2 - l^2\},$$

therefore $$z = \frac{l^2 - h^2}{4c^2} = \frac{l^2}{(y_2 + y_1)^2 - l^2}.$$

If, therefore, the mean tension at the two ends of the tape is kept constant, *i.e.* if $(y_2 + y_1)$ is constant, then z is constant, and

$$X_0 = l\left(1 - \tfrac{1}{6}z + \tfrac{3}{40}z^2 - \dots\right),$$

whence

$$X = \sqrt{l^2 - h^2}\,.\,\frac{X_0}{l} = X_0\sqrt{1 - \frac{h^2}{l^2}}.$$

To find l, put $l = X_0 + \lambda$, where λ is what is sometimes called the "sag," then

$$\lambda = l - X_0 = 2(s - x) = 2c\left(\frac{u^3}{3!} + \frac{u^5}{5!} + \frac{u^7}{7!} + \dots\right)$$

$$= 2x\left(\frac{x^2}{3!c^2} + \frac{x^4}{5!c^4} + \frac{x^6}{7!c^6} + \dots\right);$$

but $$4c^2 = (y_2 + y_1)^2 - l^2 = \frac{4T^2}{w^2} - l^2,$$

where T is the standard tension applied to the ends, hence if T is n times the weight of the tape,

$$T = nlw,$$

and
$$c^2 = \frac{T^2}{w^2}\left(1-\frac{1}{4n^2}\right),$$

and
$$\lambda = 2x\left\{\tfrac{1}{6}x^2\frac{w^2}{T^2}\left(1-\frac{1}{4n^2}\right)^{-1} + \tfrac{1}{120}x^4\frac{w^4}{T^4}\left(1-\frac{1}{4n^2}\right)^{-2}+\dots\right\}.$$

It is desirable that n should be large, never less than 10 and generally about 20. If $n=10$, the second term of the equation for λ is

$$\tfrac{1}{60}\cdot\frac{X_0^5}{2^5}\cdot\frac{1}{n^4l^4}\left(\frac{4n^2}{4n^2-1}\right)^2.$$

This is less than $X_0/18 . 10^{-6}$, similarly it is less than $X_0/306 . 10^{-6}$, when $n=20$. In such circumstances this can be neglected and we have

$$\lambda = \frac{X_0^3}{24}\cdot\frac{w^2}{T^2}\left(1-\frac{1}{4n^2}\right)^{-1}$$
$$= \frac{X_0^3}{24}\cdot\frac{w^2}{T^2}\left(1+\frac{1}{4n^2}+\frac{1}{16n^4}+\dots\right).$$

The second term of this is less than $X_0/24n^2 . 1/4n^2$, *i.e.* less than $X_0/96n^4$, and can frequently be neglected.

The formula $X=X_0\sqrt{1-h^2/l^2}$ is not in a convenient form for calculation, as the actual value of l will vary for each span measured. Call L the actual length of the tape between its zero marks at standard temperature and tension, S the nominal length of the tape, and L′ the actual length at temperature t above the standard temperature, then $X_0=S+a_0$, where a_0 is the error of the tape found when it is standardised. If α is the coefficient of expansion of the tape, we have

$$L' = L(1+\alpha t) = (X_0+\lambda)(1+\alpha t) = (S+\lambda+a_0)(1+\alpha t).$$

In the field we actually measure the difference a_1 between l and L′, hence

$$l = (S+\lambda+a_0+a_1)(1+\alpha t).$$

Here a_0 depends on the accuracy with which the tape is marked, a_1 on the accuracy with which the trestles are spaced, and αt on the temperature. With care there should be no difficulty in keeping (a_0+a_1) less than 10 mm., and with $n=10$, λ will be about $X_0/2400$, αt will never be more than 10^{-4}, consequently $(\lambda+a_0+a_1)$ should never be more than $X_0/1200$, and $(\lambda+a_0+a_1)(\alpha t)$ will never be more than $X_0 . 10^{-7}$ and usually much less. It can therefore be neglected. We may therefore put $l=S+A+\lambda$, where $A=a_0+a_1+S\alpha t$, and also $X_0=S+A$.

The formula for X now becomes

$$X=(S+A)\left\{1-\frac{h^2}{(S+A+\lambda)^2}\right\}^{\frac{1}{2}};$$

expanding we get

$$X=(S+A)\left\{1-\tfrac{1}{2}\frac{h^2}{S^2}\left(1+\frac{A+\lambda}{S}\right)^{-2} - \tfrac{1}{8}\frac{h^4}{S^4}\left(1+\frac{A+\lambda}{S}\right)^{-4}-\dots\right\}.$$

As shown above, $A+\lambda$ can without trouble be made less than $10^{-3}S$, hence if we neglect terms involving $(A+\lambda)^2/S^2$ we are neglecting a quantity less than $\frac{1}{18}S10^{-6}$ even when $h/S=1/3$, and generally very much less. For practical purposes, therefore,

$$X=(S+A)\left\{1-\tfrac{1}{2}\frac{h^2}{S^2}\left(1-2\frac{A+\lambda}{S}\right)-\tfrac{1}{8}\frac{h^4}{S^4}\left(1-4\frac{A+\lambda}{S}\right) - \tfrac{1}{16}\frac{h^6}{S^6}\left(1-6\frac{A+\lambda}{S}\right)-\dots\right\}$$

$$=S\left(1-\tfrac{1}{2}\frac{h^2}{S^2}-\tfrac{1}{8}\frac{h^4}{S^4}-\tfrac{1}{16}\frac{h^6}{S^6}-\tfrac{5}{128}\frac{h^8}{S^8}-\dots\right)$$

$$+S\frac{A+\lambda}{S}\left(\frac{h^2}{S^2}+\tfrac{1}{2}\frac{h^4}{S^4}+\tfrac{3}{8}\frac{h^6}{S^6}+\tfrac{5}{16}\frac{h^8}{S^8}+\dots\right)$$

$$+A\left(1-\tfrac{1}{2}\frac{h^2}{S^2}-\tfrac{1}{8}\frac{h^4}{S^4}-\tfrac{1}{16}\frac{h^6}{S^6}-\tfrac{5}{128}\frac{h^8}{S^8}-\dots\right)$$

$$+A\frac{A+\lambda}{S}\left(\frac{h^2}{S^2}+\tfrac{1}{2}\frac{h^4}{S^4}+\tfrac{3}{8}\frac{h^6}{S^6}+\tfrac{5}{16}\frac{h^8}{S^8}+\dots\right).$$

As $A(A+\lambda)/S$ is less than $\frac{1}{2}S10^{-6}$, it is negligible even on slopes of 1 in 2, so, neglecting this,

$$X=S-S\left(\tfrac{1}{2}\frac{h^2}{S^2}+\tfrac{1}{8}\frac{h^4}{S^4}+\tfrac{1}{16}\frac{h^6}{S^6}+\tfrac{5}{128}\frac{h^8}{S^8}+\dots\right)$$

$$+A+A\left(\tfrac{1}{2}\frac{h^2}{S^2}+\tfrac{3}{8}\frac{h^4}{S^4}+\tfrac{5}{16}\frac{h^6}{S^6}+\tfrac{35}{128}\frac{h^8}{S^8}+\dots\right)$$

$$+\lambda\left(\frac{h^2}{S^2}+\tfrac{1}{2}\frac{h^4}{S^4}+\tfrac{3}{8}\frac{h^6}{S^6}+\tfrac{5}{16}\frac{h^8}{S^8}+\dots\right),$$

and we can write simply

$$X=S+A-P+Q+R,$$

where

$$P=S\left(\tfrac{1}{2}\frac{h^2}{S^2}+\tfrac{1}{8}\frac{h^4}{S^4}+\tfrac{1}{16}\frac{h^6}{S^6}+\tfrac{5}{128}\frac{h^8}{S^8}+\dots\right),$$

which can be tabulated once for all in terms of h for any standard tape length;

$$Q=A\left(\tfrac{1}{2}\frac{h^2}{S^2}+\tfrac{3}{8}\frac{h^4}{S^4}+\tfrac{5}{16}\frac{h^6}{S^6}+\tfrac{35}{128}\frac{h^8}{S^8}+\dots\right),$$

which can be tabulated for any standard tape length in terms of h and A;

$$R=\lambda\left(\frac{h^2}{S^2}+\tfrac{1}{2}\frac{h^4}{S^4}+\tfrac{3}{8}\frac{h^6}{S^6}+\tfrac{5}{16}\frac{h^8}{S^8}+\dots\right);$$

this is always small and can be tabulated in terms of h for any given tape.

Strictly speaking, λ depends on X_0, but as X_0 will differ very slightly from S, R can be calculated on the assumption that $X_0=S$, and no appreciable error will be introduced.

Tables I., II., and III. for P, Q, and R will give an idea of the value of the corrections. They are calculated for S=24 m., T=10 kgm., w=0·01732 kgm., whence λ=1·728 mm. P is the usual correction for slope; it is always to be subtracted from the measured length of the base, and is always important. Q will vary in sign with A. and only becomes of importance on steep slopes and when the trestles are badly spaced. R is the correction for the change in shape of the catenary on a slope; it is only of importance on steep slopes, or when a low tension is being used; it is always to be added to the measured length of the base.

So far it has been assumed that the tape is inextensible, but provided that the mean tension is kept constant, and the tape is not stretched beyond its elastic limit, this assumption will produce no error. The formula developed above assumes that the mean tension at the ends is kept constant; if this be so, it can be shown that the true mean tension is sufficiently constant to make no appreciable error.

TABLE I

VALUES OF "P" IN MILLIMETRES FOR S=24 METRES, h IN METRES

h.	·00.	·01.	·02.	·03.	·04.	·05.	·06.	·07.	·08.	·09.
0·0	0·000	0·002	0·008	0·019	0·033	0·052	0·075	0·102	0·133	0·169
·1	0·208	0·252	0·300	0·352	0·408	0·468	0·533	0·602	0·675	0·752
·2	0·833	0·919	1·008	1·102	1·200	0·302	1·408	1·519	1·634	1·752
·3	1·875	2·002	2·134	2·269	2·409	2·552	2·700	2·853	3·009	3·169
·4	3·334	3·502	3·675	3·853	4·034	4·219	4·409	4·603	4·801	5·003
0·5	5·209	5·419	5·634	5·852	6·075	6·303	6·534	6·770	7·009	7·253
·6	7·501	7·752	8·009	8·270	8·535	8·803	9·076	9·353	9·635	9·921
·7	10·210	10·504	10·802	11·105	11·411	11·721	12·036	12·355	12·678	13·005
·8	13 337	13·672	14·012	14·356	14·704	15·056	15·413	15·774	16·138	16·507
·9	16·881	17·258	17·640	18·025	18·415	18·809	19·208	19·610	20·017	20·428
1·0	20·842	21·261	21·685	22·112	22·544	22·980	23·420	23·864	24·313	24·765
·1	25·222	25·682	26·147	26·617	27·090	27·567	28·049	28·535	29·025	29·520
·2	30·019	30·521	31·028	31·539	32·055	32·574	33·098	33·626	34·158	34·694
·3	35·234	35·779	36·328	36·881	37·438	37·999	38·565	39·134	39·708	40·286
·4	40·868	41·455	42·045	42·640	43·239	43·842	44·449	45·061	45·677	46·297
1·5	46·921	47·549	48·181	48·818	49·459	50·104	50·753	51·407	52·065	52·727
·6	53·393	54·063	54·737	55·416	56·099	56·786	57·477	58·173	58·872	59·576
·7	60·284	60·996	61·713	62·433	63·158	63·887	64·620	65·358	66·099	66·845
·8	67·595	68·349	69·108	69·870	70·637	71·408	72·183	72·963	73·747	74·534
·9	75·326	76·123	76·923	77·728	78·537	79·350	80·167	80·988	81·814	82·644
2·0	83·478	84·317	85·160	86·007	86·858	87·713	88·572	89·436	90·304	91·176
·1	92·052	92·932	93·817	94·706	95·599	96·496	97·398	98·303	99·213	100·128
·2	101·046	101·969	102·895	103·826	104·762	105·701	106·645	107·593	108·545	109·501
·3	110·463	111·428	112·397	113·370	114·348	115·329	116·315	117·305	118·300	119·299
·4	120·302	121·308	122·319	123·335	124·355	125·379	126·407	127·440	128·477	129·518
2·5	130·563	131·613	132·667	133·724	134·787	135·853	136·924	137·998	139·078	140·161
·6	141·249	142·341	143·437	144·537	145·641	146·750	147·863	148·980	150·102	151·228
·7	152·359	153 493	154·631	155·774	156·921	158·072	159·228	160·388	161·552	162·720
·8	163·893	165·070	166·251	167·436	168·626	169·820	171·018	172·220	173·426	174·637
·9	175·852	177·072	178·296	179·524	180·576	181·992	183·233	184·478	185·727	186·980
3·0	188·238	189·500	190·767	192·037	193·311	194·590	195·874	197·161	198·453	199·749
·1	201·050	202·354	203·663	204·977	206·294	207·616	208·942	210·272	211·608	212·947
·2	214·290	215·638	216·989	218·346	219·706	221·070	222·439	223·813	225·190	226·572
·3	227·958	229·348	230·743	232·141	233·544	234·952	236·364	237·780	239·199	240·625
·4	242·054	243·487	244·924	246·366	247·812	249·263	250·717	252·176	253·640	255·107
3·5	256·579	258·055	259·536	261·021	262·510	264·003	265·501	267·003	268·510	270·021
·6	271·536	273·055	274·579	276·107	277·639	279·176	280·717	282·262	283·812	285·366
·7	286·923	288·486	290·053	291·624	293·199	294·779	296·363	297·952	299·544	301·142
·8	302·743	304·349	305·959	307·573	309·192	310·815	312·442	314·074	315·710	317·350
·9	318·995	320·644	322·297	323·955	325·617	327·284	328·955	330·630	332·309	333·993
4·0	335·681	..	..	..	..	..	..	..	..	..

Table II

Values of "Q" in Millimetres for S=24 Metres, h in Metres and A in Millimetres

A.	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	20.
h.											
0·1	0·000	0·000	0·000	0·000	0·000	0·000	0·000	0·000	0·000	0·000	0·000
·2	·000	·000	·000	·000	·000	·000	·000	·000	·000	·000	·001
·3	·000	·000	·000	·000	·000	·000	·001	·001	·001	·001	·002
·4	·000	·000	·000	·001	·001	·001	·001	·001	·001	·001	·003
0·5	0·000	0·000	0·001	0·001	0·001	0·001	0·002	0·002	0·002	0·002	0·004
·6	·000	·001	·001	·001	·002	·002	·002	·002	·003	·003	·006
·7	·000	·001	·001	·002	·002	·003	·003	·003	·004	·004	·009
·8	·001	·001	·002	·002	·003	·003	·004	·004	·005	·006	·011
·9	·001	·001	·002	·003	·004	·004	·005	·006	·006	·007	·014
1·0	0·001	0·002	0·003	0·003	0·004	0·005	0·006	0·007	0·008	0·009	0·017
·1	·001	·002	·003	·004	·005	·006	·007	·008	·009	·011	·021
·2	·001	·003	·004	·005	·006	·008	·009	·010	·011	·013	·025
·3	·001	·003	·004	·006	·007	·009	·010	·012	·013	·015	·029
·4	·002	·003	·005	·007	·009	·010	·012	·014	·015	·017	·034
1·5	0·002	0·004	0·006	0·008	0·010	0·012	0·014	0·016	0·018	0·020	0·039
·6	·002	·004	·007	·009	·011	·013	·015	·018	·020	·022	·045
·7	·003	·005	·008	·010	·013	·015	·018	·020	·023	·025	·050
·8	·003	·006	·008	·011	·014	·017	·020	·023	·025	·028	·056
·9	·003	·006	·009	·013	·016	·019	·022	·025	·028	·031	·063
2·0	0·003	0·007	0·010	0·014	0·017	0·021	0·024	0·027	0·031	0·035	0·070
·1	·004	·008	·012	·015	·019	·023	·027	·031	·035	·039	·077
·2	·004	·008	·013	·017	·021	·025	·030	·034	·038	·042	·085
·3	·005	·009	·014	·018	·023	·028	·032	·037	·042	·046	·092
·4	·005	·010	·015	·020	·025	·030	·035	·040	·045	·050	·101
2·5	0·005	0·011	0·016	0·022	0·027	0·033	0·038	0·044	0·049	0·054	0·109
·6	·006	·012	·018	·024	·030	·036	·041	·047	·053	·059	·118
·7	·006	·013	·019	·026	·032	·038	·045	·051	·057	·064	·128
·8	·007	·014	·021	·027	·034	·041	·048	·055	·062	·069	·138
·9	·007	·015	·022	·030	·037	·044	·052	·059	·066	·074	·148
3·0	0·008	0·016	0·024	0·032	0·039	0·047	0·055	0·063	0·071	0·079	0·158
·1	·008	·017	·025	·034	·042	·051	·059	·068	·076	·084	·169
2	·009	·018	·027	·036	·045	·054	·063	·072	·081	·090	·180
·3	·010	·019	·029	·038	·048	·058	·067	·077	·086	·096	·192
·4	·010	·020	·031	·041	·051	·061	·071	·081	·092	·102	·204
3·5	0·011	0·022	0·033	0·043	0·054	0·065	0·076	0·086	0·097	0·108	0·216
·6	·011	·023	·034	·046	·057	·069	·080	·092	·103	·114	·229
·7	·012	·024	·036	·048	·060	·073	·085	·097	·109	·121	·242
·8	·013	·026	·038	·051	·064	·077	·089	·102	·115	·128	·256
·9	·013	·027	·040	·054	·067	·081	·094	·108	·121	·135	·269
4·0	0·014	0·028	0·043	0·057	0·071	0·085	0·099	0·113	0·128	0·142	0·284

Table III

Values of "R" in Millimetres for S=24 Metres, λ=1·728 mm., h in Metres

h.	·0.	·1.	·2.	·3.	·4.	·5.	·6.	·7.	·8.	·9.
0	0·000	0·000	0·000	0·000	0·000	0·001	0·001	0·001	0·002	0·002
1	·003	·004	·004	·005	·006	·007	·008	·009	·010	·011
2	·012	·013	·014	·016	·017	·019	·020	·022	·024	·025
3	·027	·029	·031	·033	·035	·037	·039	·042	·044	·046
4	·049	..	..	..	..	..	..	..	..	..

In practice the standard tension is applied at one end of the tape, and it makes a difference whether this is the higher or lower end of the tape. The error caused is made up of two parts, (1) the elastic extension of the tape, and (2) the change in the shape of the catenary. The change is small in any case, but not negligible on steep slopes. If each span is measured once with the standard tension applied at the upper end and once with the tension applied at the lower end, error (1) will obviously cancel out in the mean, and it can be shown that error (2) will also cancel out almost entirely so that the residual error is quite negligible. These errors can thus be easily avoided.

It remains to consider the accuracy attainable with this method of base measurement. The errors can be divided as follows:

(i.) *Errors in Standardising the Tapes.*—This is most important, as a proportional error in the standardisation of the tape gives the same proportional error in the final result. A probable error of about $0{\cdot}5\mu$ should be attainable.

(ii.) *Errors in the Readings.*—In taking readings on the millimetre scales an error greater than 0·2 mm. should be very rare. This will give a probable error of about 0·06 mm. for one reading of one scale. If twenty readings are taken in all, the probable error for one tape length will be about 0·02 mm. As this error is in no way systematic, the error for the whole base will only increase as the square root of the number of tape lengths, so the error for a 5 km. base measured with 24-m. tapes would only be about $0{\cdot}07\mu$.

(iii.) *Errors due to Temperature.*—In reasonable conditions the temperature of the tape should be known to within 1° C., which would give a maximum error of $0{\cdot}5\mu$ for any one span. The error in the mean should be much less than this, probably not more than a tenth.

(iv.) *Errors due to Change in the Tension.*—If the tension is applied during the measurement by an apparatus different from that used when the tape is standardised, a constant error may be introduced, which may be serious. An increase in tension acts in two ways—(*a*) by flattening the catenary, (*b*) by stretching the tape, the error caused by a change in tension δT being

$$2\lambda\frac{\delta T}{T}+\nu l\delta T,$$

where $1/\nu$ is the cross-section multiplied by Young's modulus. If $\delta T/T=0{\cdot}01$ the error due to (*a*) will be about $1{\cdot}5\mu$, and that due to (*b*) about $3{\cdot}0\mu$, ν being about 0·0007. The standard tension should therefore be kept constant to within $\frac{1}{10}$th per cent to ensure an error not greater than $0{\cdot}5\mu$.

(v.) *Errors due to Change in Weight due to Dirt, etc.*—A change in w will produce a change in the shape of the catenary, and will alter λ by an amount $2\lambda\delta w/w$, and with the standard wire when $\lambda=1{\cdot}72$ mm., $w=17{\cdot}32$ gm., a change of 1 per cent in w will cause an error of 0·035 mm., or $1{\cdot}5\mu$.

(vi.) *Errors due to Assuming Incorrect Values for* T *and* w.—Such incorrect assumptions will lead to an incorrect value for λ. If the tape has been standardised on the flat, and then used in catenary, it is necessary to know λ with the same accuracy as A, and any errors in the values assumed for T and w will produce the errors given above (iv. and v.) for a *change* in T and w. If the tape has been standardised in catenary the correction R is the only one affected, and the error introduced will always be less than $\delta\lambda$, and generally very much less. With $h/S=0{\cdot}5$, δR is about $0{\cdot}3\delta\lambda$.

$$\delta\lambda-2\lambda\frac{\delta T}{T}+2\lambda\frac{\delta w}{w},$$

hence δR is less than

$$0{\cdot}6\frac{\delta T}{T}+0{\cdot}6\frac{\delta w}{w},$$

hence an uncertainty of 1 per cent in T or w will only produce an error of about $0{\cdot}2\mu$, even on very steep slopes, when using the ordinary 24-m. wire.

(vii.) *Errors due to Inaccurate Measurement of* h.—An error in h affects all three corrections P, Q, and R, but P is the most important.

$$\delta P=\frac{h}{S}\delta h+\frac{1}{2}\frac{h^3}{S^3}\delta h+\ .\ .\ .\ ,$$

which increases rapidly as h increases. A probable error of 1·5 mm. in h is easily obtained, in this case $\delta P=0{\cdot}00006h$, if $h=2$ m. for every span the probable error of P per span is 0·12 mm., or 5μ. This error, however, will largely cancel out, and for a base of 200 spans the probable error will only be $5/\sqrt{200}$, or $0{\cdot}35\mu$; even with $h=4$ it will still be under $0{\cdot}7\mu$. With care the probable error of h can be reduced to 0·5 mm., in which case slopes of $\frac{1}{5}$ can be measured, still keeping the final probable error below 1μ.

(viii.) *Effect of Change of Gravity.*—If the tape is standardised in one place, and used at another, the effect of change of g must be considered. An increase in the value of g will increase T and w, both in the proportion $\delta g/g$. The increase in T will increase l by an amount $\nu l T\delta g/g$, and this can, if necessary, be allowed for. As w/T remains constant λ will not be appreciably affected. If the tension be applied by a spring and not by a weight, T will remain constant, but w will alter. In this case l will remain constant but λ will be slightly increased, and this can also be allowed for.

References for § (42)

For fuller details of the above theory see *Professional Papers of the Ordnance Survey*, New Series, No. 1, 1902.

"Use of Invar Wires," Benoit and Guillaume (*loc. cit.*).

(The tables for correction due to change in shape of the catenary on a slope are incorrect, being about double the correct amount.)

Report of U.S. Coast and Geodetic Survey, 1892, App. 8.

A. E. Young, *Phil. Mag.* xxix. 90 (considers the effect of the rigidity of the tape).

A paper by G. H. Knibbs (*Journal of the Royal Society of New South Wales*, xix. 20) on "A System of Accurate Measurement by Means of Long Steel Ribands," gives formulas which apply more especially to the method by which the sag correction is eliminated by altering the tension.

C. W. Adams, in a paper on "The Measurement of Distances with Long Steel Tapes," read before the Victorian Institute of Surveyors in 1888, gives tables for the sag corrections for a number of cases.

The last three papers apply more to accurate traverse work than to base measurement.

E. O. H.

General References

A. R. Clarke, *Geodesy* (London).

G. L. Hosmer, *Geodesy* (New York).

C. F. Close, *Text-book of Topographical and Geographical Surveying* (London, 2nd ed., 1913).

W. Jordan, *Handbuch der Vermessungskunde* (Stuttgart, 3 vols.).

J. Whitelaw, Jun., *Surveying as Practised by Civil Engineers and Surveyors* (London).

J. B. Johnson, *Theory and Practice of Surveying* (New York, 17th ed., 1914).

Encyklopädie der mathematischen Wissenchaften (Leipzig). (Vol. vi. deals with Surveying and gives many references.)

J. H. Gore, *A Bibliography of Geodesy*, 2nd ed., issued with the *Report of the Superintendent of the U.S. Coast and Geodetic Survey for* 1901–2, p. 429 (Washington, 1903).

The Accounts of the Ordnance Survey, The Survey of India, The U.S. Coast and Geodetic Survey, and similar publications should be referred to for details of actual large surveys.

General reports and summaries of progress appear in the *Verhandlung der allgemeine Konference der europäischen Gradmessung*, after 1886 *der Internationalen Erdmessung* (Berlin, 1865–1912).

SYMMETRY, ELEMENTS OF, BELONGING TO A CRYSTAL: axis of symmetry. See "Crystallography," § (6) (ii.).

Plane of symmetry. See *ibid.* § (6) (i.).

SYMMETRY OF CRYSTALS, CLASSES OF. See "Crystallography," § (8).

— T —

TACHOMETRIC OR STADIA SURVEYS, general methods. See "Surveying and Surveying Instruments," § (20).

TALBOT'S LAW, concerns the apparent brightness of intermittent lights. See "Eye," § (18).

TANK FURNACES FOR MELTING GLASS. See "Glass," § (15) (ii.).

TAPES IN CATENARY, THEORY OF. See "Surveying and Surveying Instruments," § (42).

TAPES OR WIRES IN CATENARY FOR BASE MEASUREMENTS. See "Surveying and Surveying Instruments," § (41) (ii.) and (iii.).

In traverses. See *ibid.* § (12).

TARTARIC ACID, OPTICAL ROTATORY POWER OF

§ (1) HISTORICAL.—The optical rotatory power of tartaric acid was discovered in 1832 by Biot,[1] who devoted one of his longest memoirs [2] to a detailed account of its properties when mixed with water, with alcohol, and with wood-spirit. Biot found that tartaric acid, when "dissolved in different fluid media, exercises on the planes of polarisation of light a special power, which distinguishes it from all other substances studied hitherto." These had agreed with quartz in obeying, at least approximately, Biot's law of inverse squares, according to which the rotation is proportional to the reciprocal of the square of the wave-length of the light.[3] This agreement had been verified in the case of turpentine, alone and mixed with ether, and of cane-sugar dissolved in water, (i.) by comparing the tints with those produced by equivalent plates of quartz, and (ii.) by eliminating the effects of rotatory polarisation with the help of a quartz plate of opposite sign acting as a compensator. When, however, tartaric acid was compared with quartz, no such parallelism was observed, the rotations of the common dextro-rotatory acid for the chief colours being as follows: [4]

	Quartz.	Tartaric Acid.
Red	18·00°	38° 7′
Orange . . .	21·40	40 29
Yellow	23·99	42 51
Green	27·86	46 11
Blue	32·31	44 40
Indigo	36·13	42 9
Violet	40·88	39 38

Similar phenomena were observed when tartaric acid was dissolved in alcohol,[5] but Biot found that "when it combines with basic substances in the same media, it loses its special action and imprints on the products the properties common to all bodies endowed with rotatory power." [6]

§ (2) INFLUENCE OF WATER.—Biot discovered that "in aqueous solutions of tartaric acid at a given temperature, the rotatory power of the acid calculated for each simple ray is always of the form $A + Be$, where e represents the proportion by weight of water in the solution." [7] This law was illustrated by plotting a against e for a series of wave-lengths, in a diagram (*Fig.* 1) which resembles the "Characteristic diagram" of Armstrong and Walker.[8] Biot's linear law is only an approximation, but it enabled him to predict that the rotatory power A of the anhydrous acid, for the red light transmitted by glass coloured with cuprous oxide, would change sign at 23° C., being positive above this temperature and negative below it.[9] This prediction was verified, after an interval of over ten years, when Laurent [10] in 1849 discovered a method by which moistened tartaric acid could be fused and cooled to a transparent glass in thicknesses up to 76 mm.; a column 70 mm. in thickness at 3·5° C. then

[1] *Mém. Acad. Sci.*, 1835, xiii., Table G, p. 168; paper read Nov. 5, 1832.
[2] *Ibid.*, 1838, xv. 93-279; paper read January 11, 1836.
[3] See "Quartz, Rotatory Power of," § (3) (i.).
[4] *Mém. Acad. Sci.*, 1838, xv. 236.
[5] *Mém. Acad. Sci.*, 1838, xv. 245.
[6] *Ibid.*, 1838, xvi. 229; paper read Nov. 27, 1837.
[7] *Ibid.* xv. 216; *Comptes Rendus*, 1835, i. 459.
[8] *Proc. Roy. Soc.*, 1913, A, 88, 388-403.
[9] *Mém. Acad. Sci.*, 1838, xvi. 269.
[10] *Ann. Chim. Phys.*, 1850, xxviii. 353.

gave $[\alpha]_{red} = -2{\cdot}787°$, where Biot's calculations gave the value $-2{\cdot}752°$. Quite recently these experiments have been extended by Bruhat,[1] who has measured the rotatory powers of the fused or glassy acid at temperatures ranging from 15° to 185° C., and for colours ranging from w.l. $430\mu\mu$ to $681\mu\mu$; these rotations are therefore available as limiting values to check the formulae used to express the influence of dilution with water on the rotatory power of the acid. Thus

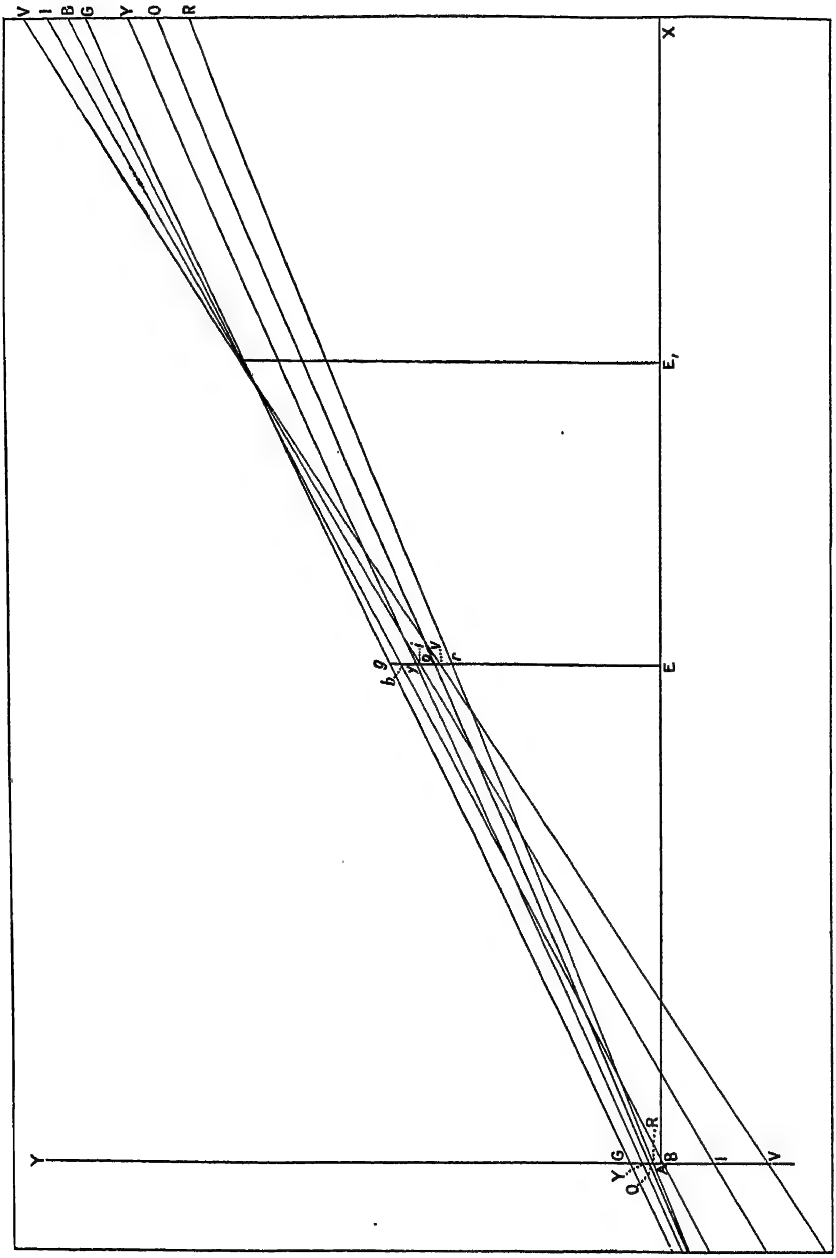


FIG. 1.

[1] *Trans. Faraday Soc.*, 1914, x. 89.

Winther[1] made use of a parabolic formula to express his measurements, which were not covered by the linear law; but this parabolic formula gives larger errors for the anhydrous acid than does the linear formula. The most recent investigations have shown that an equation with five arbitrary constants would probably be required to express accurately the relationship between rotatory power and concentration; apart from this, a linear law is probably the best approximation, *e.g.* the following formulae are correct at $e=0\cdot55$ and $0\cdot85$, and do not differ very widely from the observed values even at $e=0$.

TABLE I

SPECIFIC ROTATIONS OF TARTARIC ACID IN AQUEOUS SOLUTIONS AT 20°

Wave-length.	Rotation.
Cd 6438	$[\alpha]=\ \ 0\cdot7733+12\cdot0867e$
Na 5893	$[\alpha]=\ \ 0\cdot1100+14\cdot7433e$
Hg 5780	$[\alpha]=-\ 0\cdot1416+15\cdot4183e$
Hg 5461	$[\alpha]=-\ 1\cdot0350+17\cdot6000e$
Cd 5086	$[\alpha]=-\ 2\cdot6736\ \ 20\cdot6220e$
Cd 4800	$[\alpha]=-\ 4\cdot9528+23\cdot8233e$
Cd 4678	$[\alpha]=-\ 6\cdot3217+25\cdot5067e$
Hg 4358	$[\alpha]=-11\cdot1982+30\cdot2167e$

§ (3) ROTATORY DISPERSION IN TARTARIC ACID AND ITS ESTERS.—Although the rotatory dispersion of tartaric acid and many of its derivatives is highly anomalous, it has been found that the rotatory powers of methyl and ethyl tartrate (*Fig.* 2), both as homogeneous liquids and in a wide range of solutions, can be expressed by formulae of the type[2]

$$\alpha=\frac{k_1}{\lambda^2-\lambda_1^2}-\frac{k_2}{\lambda^2-\lambda_2^2}.$$

In this formula the positive and the negative term are *each* of the type that serves to express the rotatory power of the large array of substances which exhibit "simple" rotatory dispersion, as expressed in the formula

$$\alpha=\frac{k}{\lambda^2-\lambda_0^2}.^3$$

This result is in accordance with the view of Biot, subsequently elaborated by Arndtsen,[4] that anomalous rotatory dispersion can be produced by, and is commonly due to, the incomplete compensation of the rotations of two compounds, opposite in sign but of unequal dispersive power, so that complete compensation is only possible for one wave-length at a time.

The dispersion curves for tartaric acid resemble those for the esters very closely, *e.g.* Bruhat's curve for the specific rotatory power of glassy tartaric acid at 44·6° is

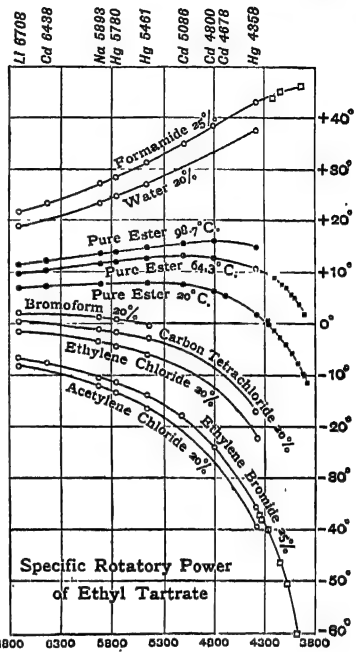


FIG. 2.

identical with, and can be superposed on, Winther's curve for methyl tartrate at 80°; there can, therefore, be no doubt that the glassy acid behaves as a lower homologue of the esters. The dispersion-curves for aqueous solutions of the acid, which are shown graphically in *Fig.* 3, can also be represented by the same formula as in the case of the esters, the values for the constants in a few typical cases being as follows:

TABLE II

ROTATORY DISPERSION IN AQUEOUS SOLUTIONS OF TARTARIC ACID AT 20°

	$\alpha=\frac{k_1}{\lambda_2^2-\lambda_1^2}-\frac{k_2}{\lambda^2-\lambda_2^2}$.			
	λ_1^2.	λ_2^2.	k_1.	k_2.
0·45	0·030	0·074	17·127	12·093
0·50	,,	,,	17·485	12·043
0·55	,,	,,	17·686	11·877
0·60	,,	,,	18·053	11·865
0·65	,,	,,	18·367	11·812
0·70	,,	,,	18·709	11·799
0·75	,,	,,	18·936	11·714
0·80	,,	,,	19·160	11·624
0·85	,,	,,	19·485	11·605
0·90	,,	,,	19·657	11·475
0·95	,,	,,	19·592	11·108

[1] *Zeitschr. physikal. Chem.*, 1902, xli. 186.
[2] Lowry and Dickson, *Trans. Chem. Soc.*, 1915, cvii. 1173; Lowry and Abram, *ibid.* p. 1187.
[3] Lowry and Dickson, *ibid.*, 1913, ciii. 1067.
[4] *Ann. Chim. Phys.*, 1858, liv. 421.

§ (4) Other Derivatives of Tartaric Acid.—In the case of the tartrates of the alkali metals, the principal anomalies noted in the acid disappear, as Biot recorded in 1837; thus the solutions are consistently dextrorotatory and the rotations increase progressively with diminishing wave-length (*Fig.* 4), in some cases in close approximation to the law of inverse squares. A careful study of these rotations[1] has shown, however, that they cannot be expressed by the one-term equation which applies to cases of "simple" rotatory dispersion, but require a formula with two terms, just as in the case of the acid; the negative term is, however, very much smaller, so that the rotations are positive throughout the visible and part of the ultra-violet spectrum. The most interesting of the derivatives of tartaric acid are boro-tartaric acid and tartar emetic, since these alone amongst the substances hitherto investigated give dextrorotations obeying the "simple" dispersion law. In the case of tartar emetic the rotations become negative when the salt is dissolved in concentrated potassium hydroxide, but these solutions, like those of the original salt, exhibit simple rotatory dispersion.

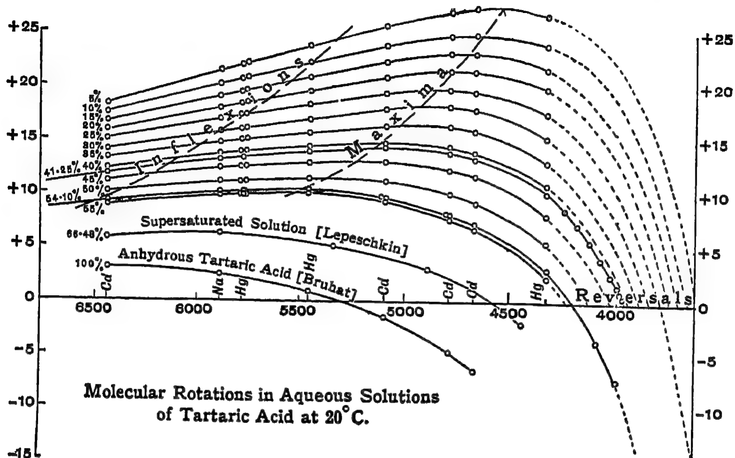


Fig. 3.

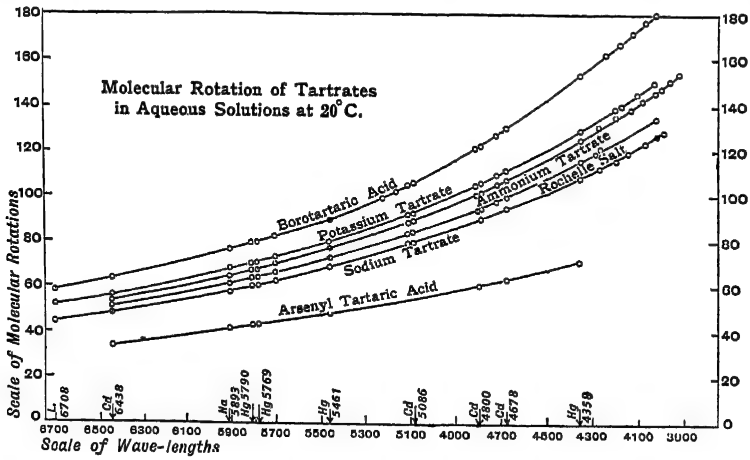


Fig. 4.

§ (5) Origin of Anomalous Rotatory Dispersion.—The vast majority of optically-active compounds, including the secondary

[1] Lowry and Austin, Bakerian Lecture, 1921.

alcohols, and the sugars, exhibit simple rotatory dispersion, although they may contain a considerable number of asymmetric carbon atoms, often in close association with the unsaturated linkages to which so many optical anomalies are due. The structural formula $HO \cdot CO \cdot CHOH \cdot CHOH \cdot CO \cdot OH$ commonly assigned to tartaric acid gives no hint as to why this acid and its esters should give anomalous rotatory dispersion, when the closely related sugars do not. It is therefore probable that this formula is not a complete representation of the molecular structure of the acid as it exists in the fused state or in solution, although it may perhaps provide a correct picture of the structure of the solid crystalline acid. In the case of nitrocamphor [1] it has been proved that the solutions contain a labile isomeride of opposite rotatory power, which is reconverted into the original compound when the solutions are allowed to crystallise. This labile isomeride has not been isolated, although several of its derivatives are known; these possess a high dextrorotation, whereas the parent substance is laevorotatory in most solvents. A satisfactory explanation of all the observations is given by assuming that in solution a reversible isomeric change takes place, as represented by the scheme

$$C_8H_{14}\Big\langle\begin{matrix} CH\cdot NO_2 \\ | \\ CO \end{matrix} \rightleftarrows C_8H_{14}\Big\langle\begin{matrix} C:NO\cdot OH \\ | \\ CO \end{matrix}$$

Nitrocamphor. Pseudonitrocamphor.

This phenomenon is usually described as "dynamic isomerism."

In the case of tartaric acid the evidence for the existence of a labile isomeride in solution is less complete, since it is not possible to follow the progress of the isomeric change by means of the changing rotatory power or "mutarotation" of the freshly dissolved material; but the solutions have all the optical properties of a mixture of isomerides of opposite rotatory power and unequal dispersion, and there are no observations which contradict the view that the hypothesis of Biot and Arndtsen as to the origin of the anomalous dispersion can be applied.

The structure to be assigned to the labile form of tartaric acid has not yet been determined, although derivatives of both forms are now known (see § (4) above) which exhibit simple rotatory dispersions of opposite sign and unequal magnitude.

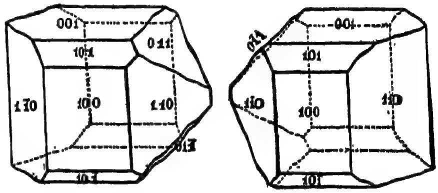
§ (6) DEXTRO- AND LAEVO-TARTARIC ACID. —In virtue of the two asymmetric carbon atoms which it contains, tartaric acid can exist in four modifications, of which two are optically active and two inactive. These may be represented conventionally thus:

$$\underbrace{\begin{matrix} HO\cdot CH\cdot COOH \\ | \\ HOOC\cdot CH\cdot OH \\ d\text{-Tartaric acid.} \end{matrix} \qquad \begin{matrix} HOOC\cdot CH\cdot OH \\ | \\ HO\cdot CH\cdot COOH \\ l\text{-Tartaric acid.} \end{matrix}}_{\text{Racemic acid (inactive).}}$$

$$\begin{matrix} HO\cdot CH\cdot COOH \\ | \\ HO\cdot CH\cdot COOH \end{matrix}$$

meso-Tartaric acid (inactive).

Racemic acid was discovered by Kestner in 1822, and regarded as an "isomer" of the common tartaric acid, which Scheele had described in 1769; but it was not until 1848–53 that Pasteur, by carefully sorting out the crystals of sodium ammonium racemate, proved that racemic acid is a mixture of the common dextrorotatory tartaric acid with an



Dextro-tartaric Acid. Laevo-tartaric Acid.

FIG. 5.

equal amount of an enantiomorphous laevorotatory acid. The properties of this acid are an exact replica of those of the dextro-acid except as regards rotatory power, crystalline form, etc., where a reversal of sign is possible. All the phenomena of anomalous rotatory dispersion, which have been described in the previous paragraphs for dextro-tartaric acid, are therefore repeated in laevo-tartaric acid, but with the sign of the rotations reversed.

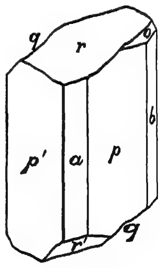


FIG. 6.

The relationship between rotatory power and crystalline form which exists in the case of quartz (*q.v.*) is found again in tartaric acid and the tartrates. Thus the crystalline forms of *d*- and *l*-tartaric acid, as shown in *Fig.* 5, are enantiomorphous just like dextro- and laevo-quartz. On mixing these two acids, however, they unite to form a racemic acid (*Fig.* 6) which crystallises in a form in which enantiomorphism does not occur. All the

[1] Lowry, *Trans. Chem. Soc.*, 1899, lxxv. 211.

metallic tartrates unite in a similar manner with their isomerides, giving rise to inactive double-salts or "racemates" in which no enantiomorphism exists. In one solitary instance, however, namely, in the case of sodium ammonium tartrate, the racemate breaks up, at temperatures below 26°, into the two component tartrates, which therefore separate from cold solutions in enantiomorphous crystals of equal and opposite rotatory power; it was by the study of this unique case that Pasteur was led to the discovery of the relationship between optical activity and crystalline structure and to the experiments which culminated in the isolation of laevo-tartaric acid. T. M. L.

Taylor, H. Dennis, details of triple astronomical objective by. See "Telescope," § (5).

Teeth-cutting for Divided Circles. See "Divided Circles," § (8).

Telemeter, Stereoscopic. See "Range-finder, Short-base," § (2).

Telephone as Sound Reproducer. See "Sound," § (58).

Telephotometer: an instrument for measuring the brightness of a distant surface. See "Photometry and Illumination," § (118).

TELESCOPE, THE

The aim of this article is to describe the telescope as a physical instrument, defined by the limitations which are imposed upon it by the purposes to which it is directed, by the construction of its optical parts, and by the geometrical features of its support or mounting. This will exclude the description of some forms, the interest in the construction of which is purely astronomical or which are otherwise restricted to special technical application. It will limit the theory to those geometrical cases that subserve the purpose of rendering visible inaccessible and distant objects. In essence, the method of making a distant object accessible is to replace it by the image of it which a converging lens or mirror forms near its principal focus. This image may then be impressed upon a photographic plate, or examined visually with a second lens. In the latter case the second lens or eyepiece is treated as part of the telescope. This difference separates the two cases so widely, as problems of optics, that they will require in large measure separate theoretical treatment, different features rising into importance or dropping out of it, in the one case or the other.

§ (1) Geometrical Theory.—We shall begin by rehearsing the results of geometrical optics as far as they apply to the visual telescope consisting of an objective and eyepiece. The standard case—to which any other actual ones are approximate—for an object given "at infinity," places the image also "at infinity," since the normal eye at rest brings a parallel beam to a focus on the retina. Hence, regarded as a general optical instrument, the telescope is a degenerate case, with all the cardinal points at infinity and the focal length infinite also. Only one case need be considered, namely, that in which objective and eyepiece both act like converging lenses. The Galilean combination, in which the eyepiece is a negative lens, is now obsolete even as a field glass, though it has found application elsewhere in the telephoto lenses of cameras. In regard to eyepieces it is customary to distinguish one of the two standard forms, the Ramsden, as positive, and the other, the Huyghens, as negative. But in theory both are of the same type, acting like a single converging lens, in which the anterior principal focus precedes the anterior unit point, the essential difference being that for the Huyghens eyepiece this principal focus lies between the two lenses and is therefore inaccessible for spider threads or the apparatus of measurement. In fact the Huyghens eyepiece, turned round, is used as a magnifying lens under the name of Wollaston's doublet. See below, § (6).

Moreover, as far as simple point-to-point correspondence goes, any combination of lenses used as objective or as eyepiece differs from the ideal "thin lens" only in separating the two unit points. Suppressing this difference, we may summarise all the cases under a single model.

Let A, a (*Fig.* 1) be the positions of the objective and eye lenses respectively, or of

Fig. 1.

the (united) unit points of the objective- and eye-combinations, the lenses being treated as thin; F, F′, f, f' the anterior and posterior principal foci of these. In the present case F′ and f coincide. Let $\Phi = \mathrm{FA} = \mathrm{AF}'$, $\phi = fa = af'$. Now any ray entering parallel to the axis will emerge parallel to the axis, its distance from the axis being reduced by the factor $af/\mathrm{AF}' = -\phi/\Phi = -1/m$, say. It follows that if Qq is the image of any object Pp, the linear magnification is $-1/m$, the same for all distances. To find the relative

positions of P, Q, call R the intermediate focus. Then[1]

$$\frac{\Phi^2}{\mathrm{FP}}=\mathrm{RF}'=\mathrm{R}f=\frac{\phi^2}{f'\mathrm{Q}},$$

or $$\mathrm{FP}=m^2 . f'\mathrm{Q}.$$

If P is to the left of F, Q is to the left of f'. Equally, if P_1, P_2 be any two objects and Q_1, Q_2 their images,

$$P_1P_2=m^2 . Q_1Q_2.$$

It is to be noted that F, f' are themselves a pair of conjugate points. So also are points at infinity.

If M be the object which gives an image at a, and N the image of an object at A, we have

$$\mathrm{MA}=m . \mathrm{A}a=m^2 . a\mathrm{N},$$

and the object space and image space are divided up into corresponding regions as follows (*Fig.* 2):

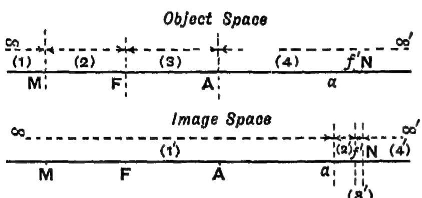


Fig. 2.

For rays not parallel to the axis, the construction is obvious. If y, y' are the relative breadths of object and image, and β, β' the angles between original rays 1, 2 and the corresponding emergent rays 1′, 2′, we have (*Fig.* 3)

Fig. 3.

$$y=\beta . P_1P_2,\ y'=\beta' . Q_1Q_2;$$

but $$P_1P_2=m^2 . Q_1Q_2$$

and $$y=-m . y',$$

so that $$\beta'=-m\beta,$$

and the magnification of angle between any two rays is the same for all positions, as, of course, follows from the general theorem that the angular magnification is the reciprocal of the linear magnification. The apparent distance for any object is simply $1/m$ times the actual distance.

[1] See "Lenses, Theory of Simple," § (3).

The field of view is governed by the apertures $2b$, $2B$ of the eye glass and objective.

The object glass gives as image (*Fig.* 4) a ring of radius $NN'=B/m$, which is the "exit-pupil," through which all rays meeting the

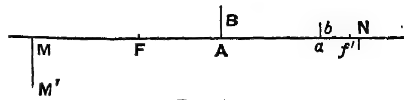


Fig. 4.

object glass pass on emergence. Similarly, the eye glass is the image of a ring $M'=bm$, and this ring is the "entrance-pupil," through which, again, all incident rays must pass. It follows that the field of view which receives full illumination is of angular radius $(bm-B)/MA=(b-B/m)/Aa$, while the radius of partial illumination extends to $(b+B/m)/Aa$. It will be remarked as evident that, apart from considerations of full and partial illumination by the object glass, the field commanded by the eyepiece is simply b/Aa, and if $b/B=m=\phi/\Phi$, no beams except those parallel to the axis will pass completely through the instrument.

By an increase of the aperture the fully illuminated field is diminished, while the partially illuminated field is increased.

Any beam of light filling the object glass of radius B, and emerging through the exit-pupil of radius, say, p_0, is thereby intensified in the ratio $(B/p_0)^2$, or m^2. But if the beam issues from an object of finite area, it will be spread on emergence over an image increased in angular area in the same ratio, and the "brightness of the image will be equal to the brightness of the object," except for loss by absorption and reflection of light in the telescope. To see the significance of this statement clearly, the part played by the eye must be considered.

In *Fig.* 5 Pp is an object; two rays PA, pA meeting the object glass at A emerge as QN, qN. If the eye be placed so that N is its ante-nodal

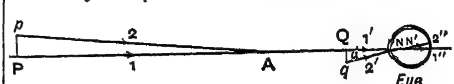


Fig. 5.

point, these rays proceed in parallel directions through N′ the post-nodal point, and the area of the retina covered by the image is seen to be $m^2\times$ that which would have been covered had the telescope been removed and the eye placed at A. Two qualifications, however, fall to be noticed. The retina is not continuous, and fails to recognise stimuli as distinct which fall, it is supposed, on the same element of its structure. These elements subtend angles of about 1 minute of arc at N′. Hence, if Qq

subtends less than this angle at N, the intensification of the beam which the object glass effects will not be dissipated in its effect on the retina, and the brightness of a "point-source" will be increased by a telescope in the ratio m^2. Or, rather, in comparing the light gathered by the telescope with what the eye unaided would acquire, we have the ratio B^2/p^2, where p is the radius of the pupil of the eye, and not m^2, which is equal to $B^2/p_0{}^2$. Hence, for a telescope applied to a point-source, the intensification of light is constant and not dependent on the magnification employed, apart from changes in the pupil of the eye; but if applied to a finite area the intensity is never increased, and when the magnification is sufficiently high it is diminished in the ratio $(p_0/p)^2$. Thus, for example, taking the diameter of the pupil of the eye as one-fifth of an inch, with a five-inch telescope the intensification of the light of a star would be about 25^2 times, equal to 7 magnitudes, while the background of the sky would be progressively darkened, if the eyepiece employed gave a magnification above 25 times; a power of 100 would darken it 16-fold.[1]

§ (2) Defects of the Simple Theory.—This is the primitive theory of the telescope. We must now indicate how it is modified by actual circumstances, namely, (1) aberrations, or faults in the theory of strict linear correspondence[2] in the object and image space; (2) diffraction phenomena; and (3) chromatic qualities of the glass employed, and other considerations.

In dealing with these points we may now pass also from the visual telescope to the photographic instrument and direct attention to the faults of the field produced by the objective at its principal focal plane, treating the eyepiece briefly afterwards from the point of view of its capacity to correct some of the faults.

The telescope is distinguished from other image-forming systems as an instrument in which the ratio of semi-aperture to focal length is small, seldom reaching the value 1/10. The angular radius of the field considered is of the same order. These are the elements upon which the geometrical aberrations depend, with numerical factors arising from the curvatures of the refracting or reflecting surfaces, their separations, refractive indices, etc. In such a case the aberrations can be expressed in a series proceeding by ascending odd powers of the aperture ratio and the angular radius of the field.

Thus, if the exact expression for size and position were worked out, for the image corresponding to any given object, if we retain only terms of the first order of the angular radius, we have the exact linear correspondence of object and image field sketched above; but if we retain the third order of both the variables jointly, we have a treatment of aberrations sufficient for the telescope.

§ (3) Von Seidel's Five Aberrations.—Under the above limitation the independent aberrations are five in number only, for all pairs of conjugate foci, as was shown by Seidel.[3] Supposing only one is present at a time, they permit of simple description. Using B, Φ as above, so that 2B/Φ is the aperture ratio, let β be the angular distance of the object point from the axis of the telescope. Let $\delta_1 G$, etc., represent the aberrational coefficients expressed in terms of the curvatures and refractive indices of the lenses.[4] Then we can describe the matter as follows:

(i.) *Spherical Aberrations.*—Rays which are originally parallel to the axis, from different zones of the object glass, strike the axis at different points (*Fig.* 6). It follows that the whole beam is collected within a "least circle

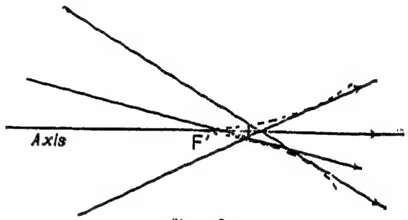


Fig. 6.

of aberration" at a certain distance from the principal focus. The angular radius of this circle can be shown to be $\frac{1}{8}B^3/\Phi \,.\, \delta_1 G$, and it comes after the principal focus by a distance $\frac{3}{8}B^2\Phi \,.\, \delta_1 G$. The extreme aberration is $\frac{1}{2}B^2\Phi \,.\, \delta_1 G$.

(ii.) *Coma.*—For an oblique ray, for the same zone, different points do not bring to coincidence parallel rays impinging on them. Rays from diametrically opposite points of the zone intersect, but the intersections are distributed round a circle (*Fig.* 7), the centre of which is displaced, by an amount equal to its diameter, from the focus corresponding to the middle of the object glass; in consequence the light from a point is distributed in a flare or fan, of 60° opening, with its tip at the linear focus. The "comatic radius" is $\frac{1}{2}\beta \,.\, B^2/\Phi \,.\, \delta_2 G$. The condition $\delta_2 G = 0$, implying equality of magnification for different zones, is sometimes known as the Fraunhofer condition, and sometimes as Abbe's sine condition, from the form it takes for wide angle systems,

[1] On the use of night glasses, for terrestrial objects, cf. Rayleigh, *Collected Works*, ii., Art. 82, 96.
[2] For further details see "Optical Calculations," § (2).

[3] *Ast. Nach.*, 1853, xxxvii.
[4] Methods of calculating these coefficients will be found in "Optical Calculations."

for which naturally it is of supreme importance. It is of hardly less importance for astronomical photography, and in the Newtonian reflector with the high aperture-ratios employed, the presence of coma spoils the images and scatters the light most seriously when the field is wide.

(iii.) and (iv.) *Astigmatism and Curvature of the Field.*—These are faults depending upon the square of angular radius of the field. Rays from different points of the objective are not

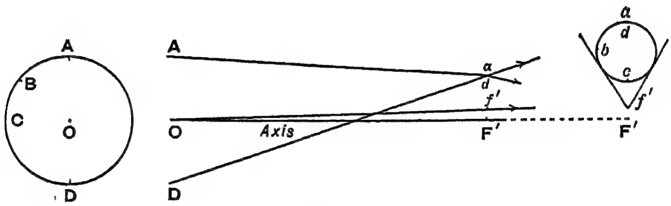


FIG. 7.

brought to any one focus. They do, however, all meet two "focal lines," a secondary focal line in the plane of the axis, and a primary perpendicular to it (*Fig.* 8), while midway between the two lines they all pass within a certain "focal circle," which gives the best approach to stigmatic representation. The curvature of the field may be considered to be that of the surface upon which the focal circles lie. The radius of the focal circle is $\frac{1}{2}\beta^2 . B/\Phi . \delta_2H$, and

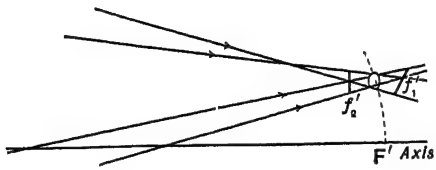


FIG. 8.

the curvature of the field $1/\Phi . (\delta_3G+\delta_2H)$,—two independent quantities. The condition which is somewhat erroneously known as Petzval's "condition for flatness of field" implies $\delta_3G=-\delta_2H$, and only gives a flat field if $\delta_2H=0$ or if the system is otherwise made stigmatic.

(v.) *Distortion.*—This is a representation of the object field on the image field stigmatically, but on a scale that varies with the angular radius β. The angular measure of distortional displacement is denoted by $\beta^3/F . \delta_3H$.[1]

As a simple example we may write down the values of these coefficients for the concave mirror of radius R, and therefore focal length R/2. We have $\delta_1G=-(1-e)/R^2$, $\delta_2G=-1/R$, $\delta_3G=0$, $\delta_2H=-1$, $\delta_3H=0$, where e measures the departure of the figure from a spherical shape, $e=1$ corresponding to a parabolic section. The distortion is zero. The comatic displacement is $2\beta . B^2/R^2$, directed outwards; for example, with an aperture $\Phi/5$, this reaches 10″ at a distance of 34′ from the centre of the field; for the same figure the radius of the least circle of aberration is $(1-e)\times 6''\cdot 4$.

For a system of two thin lenses in contact, the values of the coefficients may be written down in terms of curvatures and refractive indices, and furnish a good guide to the actual case of an ordinary objective, since the aberrations are not greatly altered by small separations between the surfaces. Denote by C_1, C_2 the curvatures of the anterior and posterior faces of the first lens, reckoned positive when convex to the ray, $1/n$ its refractive index, and write $k=(1-1/n)(C_1-C_2)$, $\mathfrak{p}=kn$, $q=(1+1/n)(C_1+C_2)$, so that k is the reciprocal of the focal length, $\mathfrak{p}$ is Petzval's expression, and q measures the distribution of curvature between the faces for given focal length, and distinguish the like quantities for the second lens by an accent; then, omitting δ_1G, the expression for which is too complicated to be of much service, we have

$$k+k'=K=-\frac{1}{\Phi},$$

$$\begin{aligned}K\delta_2G &= -\tfrac{1}{2}(nkq+n'k'q') \\ &\quad -\tfrac{1}{2}nk^2-n'kk'-\tfrac{1}{2}n'k'^2-(k+k')^2 \\ &= -\tfrac{1}{2}(nkq+n'k'q') \\ &\quad -\tfrac{1}{2}(n-n')k^2-(1+\tfrac{1}{2}n')(k+k')^2,\end{aligned}$$

$$\delta_3G=-\left(1+\frac{\mathfrak{P}}{K}\right) \qquad \mathfrak{P}=\mathfrak{p}+\mathfrak{p}'$$

$$\delta_2H=-1,$$

$$\delta_3H=0.$$

Hence for a pair of thin lenses in contact the distortion is zero, the radius of the focal circle depends only on the aperture ratio B/F, and not at all upon the curvatures of the faces; the same is approximately true of the curvature of the field. The comatic displacement, $2\beta . B^2/F . \delta_2G=2\beta B^2K\delta_2G$, gives an outward directed fan when positive, and inward when negative. Examining the expression we

[1] For further particulars on these coefficients of aberration see Sampson, *Phil. Trans.*, 1912, ccxii. 149; 1913, ccxiii. 27.

see that it cannot be zero unless one at least of the quantities kq, $k'q'$ is negative, corresponding to $C_1^2 < C_2^2$, $C_1'^2 < C_2'^2$, or the anterior faces flatter. Roundly speaking, a lens bulged outward towards the incident light gives coma directed towards the axis, and conversely.

Since little or nothing can be done with astigmatism and curvature of the field, these are usually passed over, and the faults of the objective considered[1] under the heads: (1) the difference of points of concurrence of rays originally parallel to the axis, according as they strike the object glass centrally or at the rim, *i.e.* a measure of the extreme spherical aberration; and (2) the difference of focal lengths for rays falling in these manners; this is equivalent to a determination of the comatic displacement as a difference of magnification, zone by zone. There are further considerations of the chromatic changes in these aberrations, namely, in the position of the principal focus for paraxial rays, in the principal focal length, and the chromatic difference of spherical aberration. The correction of this last fault is known as Gauss's Condition. The questions of chromatic correction are considered below.

§ (4) Practical Treatment of the Aberrations.—For visual work and for photographic work confined to the centre of the field, the presence of a moderate degree of coma would matter little, did it not make the definition sensitive to small faults in "squaring-on" the objective. For small cemented objectives a more common and often a gross fault is error of centring. Each lens surface possesses a definite axis and "centre," being the line joining the centres of the anterior and posterior spherical surfaces and the points where this line meets the surfaces respectively. For the two constituents of a doublet the centres of contiguous faces must be set in coincidence. The further condition that the directions of the axes must also coincide will then provide for itself if the inner surfaces are of equal radius and in contact. If such a fault is present the definition is deteriorated generally, the best being located off the axis, and if the objective is rotated in its cell, the image usually describes a circle in the field. For larger glasses, in which proper care has been exercised in the manufacture, this fault is not so much to be feared.

It must be recognised that refined optical theory has not played a large part in the development of the great astronomical object glasses. The doublet, with small ratio of aperture to focal length, is the simplest optical combination. Mathematical optics makes its performance and possibilities intelligible, but their most appropriate field of application lies in the complicated, wide-angle systems required in the camera and the microscope. The most celebrated makers of lenses have been men not much versed in theory, or have openly said it was of little use to them. Fraunhofer may be admitted as an exception. His successor Merz was the greatest maker of object glasses in the middle of last century. In particular he made a 15-inch lens for Harvard College Observatory. William Simms was engaged to examine this lens for purchase, and has left an interesting report[2] upon it, in which he says: "Mr. Merz's means of proving his object glasses are not such as appear to me best qualified to lead to a correct result. . . . It is merely the examination of a printed paper, not by directing them to the heavens, which in all cases, where it is practicable, is certainly the most desirable." Neither Alvan Clark nor Thomas Cooke was a man of theoretical training; they were rather consummate artists and craftsmen. G. Calver is of the same type. An account by Grubb[3] of the making and testing of objectives must be read in relation to its date, but it expresses the view referred to with great emphasis:

> *Object glasses cannot be made on paper.* When I tell you that a sensible difference in correction for spherical aberration can be made by half an hour's polishing, . . . you will see that it is practically not necessary to enter upon any calculation for spherical aberration. We know about what form gives an approximate correction; we adhere nearly to that, and the rest is done by figuring of the surface. To illustrate what I mean. I would be quite willing to undertake to alter the crown or flint lens of any of my objectives by a very large quantity, increasing one and decreasing the other so as to still satisfy the conditions of achromatism, but introducing theoretically a large amount of positive or negative spherical aberration, and yet to make out of the altered lens an object glass perfectly corrected for spherical aberration.

It would be idle to disregard these examples, though it may be thought that they are bad models to follow, and that theory must finally justify itself. In fact it has already done so, in the case of the two objectives of H. D. Taylor referred to in § (5) (iv.). The much simpler case of making a mirror, with its proper parabolic figure, used to be in the same position; in consequence the older mirrors are systematically deeply over-corrected; but the development of the knife-edge method of measuring the figure in different zones[4] has made the realisation of geometrical correctness of form a certainty.

The great difficulty in testing a lens arises from its convex surfaces. The spherometer is

[1] Steinheil and Voit, *Handbuch d. angewandten Optik*, i., Leipsic, 1891. (Translation by J. W. French. Blackie & Son, 1918.)

[2] *Harvard Annals*, 1846, I. p. cx.

[3] Royal Institution, 1886.

[4] Cf. C. D. P. Davies, *Mon. Not.* lxix. 355, and below, § (11). See also "Objectives, Testing of Compound," § (3) (ii.).

in the first place not delicate enough for optical work, and in the second, occupying a great deal of time, only compares with one another a number of separate points of the surface. The lens is therefore put together and examined as a whole, an artificial star near the eye end sending its light through the lens to a plane mirror which reflects it back again through the lens, forming an image that is examined close to the principal focus. The error in performance of any spot of the objective is here doubled, and may be examined with some facility, and though it is not possible to say which of the convex surfaces contributes the fault, it is immaterial, since the fault can be removed by figuring the anterior surface as the most convenient. If local variations of refractive index of the glasses occur, they might equally be dealt with by local figuring. Given sufficient skill, with the aid of these tests, lenses can be made to perform up to the theoretical limit of resolution imposed by the wave structure of light (see § (7) below).

§ (5) DETAILS OF LENSES. (i.) *Alvan Clark.*—Perhaps owing to the tentative character of the work of the great artists, there are not many numerical particulars available of the most celebrated lenses. Alvan Clark's two greatest lenses, the Lick telescope (36 inches) and the Yerkes telescope (40 inches), are similar in character. The curves of the former are given in the *Lick Observatory Publications*, i. 61, as follows (adjusting the signs to the usual convention, + for a surface convex to the incident ray):

$$R_0 = +259\cdot5 \text{ inches} = -R_2.$$
$$R_4 = -239\cdot6 \text{ inches}, \quad R_6 = +40{,}000 \text{ inches}.$$

The surfaces R_4, R_6 are separated by 6·5 inches. The refractive indices are not stated. The glass was by Feil of Paris. These curves are in general agreement with those given by Steinheil and Voit[1] as representative of the great American objectives, except that these authors make R_1 somewhat greater than R_2, and the concavity of the fourth surface more pronounced; the calculated aberrations they give show a fine performance, a minute residual coma and minute difference of focal length for colour being the only faults that remain. The separation of the lenses helps to satisfy the Gauss condition.

(ii.) *Thomas Cooke.*—Thomas Cooke worked from curves in which, approximately, for refractive indices for the D ray 1·518 and 1·620 respectively,

$$R_0 = 2\cdot000, \quad R_2 = -3\cdot000, \text{ double convex.}$$
$$R_4 = -2\cdot815, \quad R_6 \text{ large}, \quad \text{double concave.}$$

The second and third surface are contiguous. These leave a certain amount of residual inward directed coma, but are found so sound and convenient in practice that the firm he founded still adheres to them. The correction for spherical aberration is made by the anterior surface of the flint lens almost wholly, leaving the fourth surface to be modified at will to meet faults of achromatism.

(iii.) *Fraunhofer.*—The correction for coma requires the crown lens to be somewhat flatter in its anterior surface, and the flint lens to be a meniscus. The associated condition, known usually as the Sine Condition, Seidel called by Fraunhofer's name, because it is nearly satisfied in a celebrated Königsberg objective of 6·2 inches, by Fraunhofer. The radii for this lens are, to an arbitrary unit,

$$R_0 = +\cdot838, \quad R_2 = -\ \cdot334, \quad \mu_1 = 1\cdot529.$$
$$R_4 = -\cdot341, \quad R_6 = -1\cdot173, \quad \mu_3 = 1\cdot639.$$

These leave a residuum of outward directed coma. To satisfy completely the conditions for freedom from spherical aberration and coma, Steinheil[2] showed they should run

$$R_0 = +\cdot696, \quad R_2 = -\ \cdot363.$$
$$R_4 = -\cdot372, \quad R_6 = -1\cdot650.$$

Comparison between the numbers indicates clearly enough how correct was Fraunhofer's appreciation of what was required.

(iv.) *Taylor's Triple Objective.*—The difficulties of constructing a large objective are so great that simplicity of design is the dominating consideration. But there are two triple combinations, both due to H. D. Taylor, that have established themselves in astronomical work in spite of this. The first (*Fig.* 9) is the photo-visual object glass. By using three lenses and choosing suitable glasses the combination is rendered perfectly achromatic, and is referred to further below, on the question of achromatism. But the extra surfaces permit more choice in the curves, while securing absence of spherical aberration and coma. Thus the third, fourth, and last surfaces are concave and may be tested by reflection; the second radius is exactly equal to the third, and the fifth to the fourth, permitting their truth to be tested also one by one. Only the first surface remains to be tested in the combined system. The utmost skill and precaution are of course required in centring such a system. It has been found practicable to make them with an aperture F/18, without introducing undesirably large angles of incidence on the inner surfaces. The other triple system is the well-known Taylor

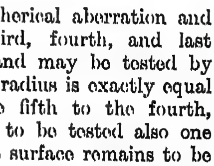


FIG. 9.

[1] *Angewandte Optik*, p. 179. Case 8a*.

[2] *Sitzungsber. k. Bayer. Akad.*, 1889, xix. 413.

photographic lens (*Fig.* 10). A description is given of one of 10 inches aperture and 45 inches focal length, constructed for Mr. Franklin Adams, in *Monthly Notices*, lxiv. 613, but one has been constructed recently for Mount Wilson Observatory to work up to an aperture F/2. The objective consists of three separated lenses, the first and third of dense barium crown glass and the middle one of flint glass. The crown lenses are double convex, their outer surfaces being the more curved; the first lens is of lower power than the third, and the distance from the first lens to the second is about half as great again as from the second to the third. Separation of the lenses permits astigmatism and curvature of the field to be corrected, coma is also corrected, and for a sufficient range of colour, spherical aberration is corrected for the neighbourhood of the ordinary photographic region of the spectrum, so that the images of stars are truly stigmatic up to a radius of $7\frac{1}{2}°$ from the axis.

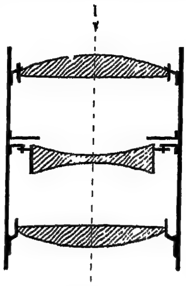


FIG. 10.

§ (6) THE EYEPIECE. (i.) *Huyghens.*—In conjunction with the optical performance of the objective should be considered, for visual use, the eyepiece. The traditional Huyghenian is still made, without much variation, and proves quite satisfactory for cases where micrometer wires or other focal scale is not required (see *Fig.* 11).

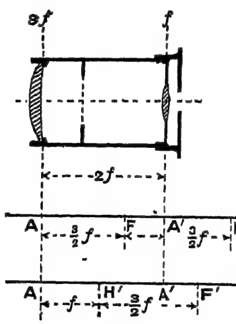


FIG. 11.

The principal focus corresponding to an emergent parallel beam lies within the combination, three-fourths of the way from the field lens to the eye lens, and the power is twice that of the field lens and two-thirds that of the eye lens. The term "negative" applied to this eyepiece is a misnomer, as remarked already, the order of the cardinal points being the same as for a converging lens. In fact, reversed, the combination is known as a magnifying lens, under the name of the Wollaston Doublet, an object at Φ′ being inspected by an eye placed beyond A.

(ii.) *Ramsden.*—With less reason the traditional Ramsden eyepiece is also preserved for micrometer work (*Fig.* 12). Its prominent defect is that particles of dust upon the field lens are visible with the eye lens, as blurs and smudges badly out of focus. A much better form, though it sacrifices breadth of the field of view somewhat, is a single triple cemented achromatic lens, as recommended by Steinheil (*Fig.* 13), which besides gives a flatter field and ample clearance from the micrometer wires. Owing to the fact that in visual work the object is invariably examined in the middle of the field, lateral aberrational or chromatic faults are of minor importance. But it is possible in some measure to compensate the inevitable curvature of field and astigmatism of the objective with those of the eyepiece, especially with the Steinheil form, which leaves some liberty in construction. The same is true of colour, but as the eye itself is not strictly achromatic, the discussion is rather academic than useful, because eyepiece and eye together should be regarded as a united combination for conveying to the retina the image formed in the focal plane of the objective, and in practice the eye is used in such a way as to minimise faults that are present.

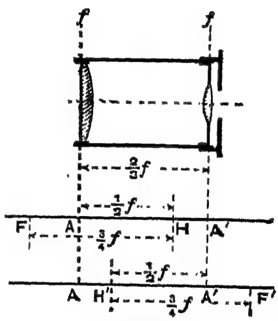


FIG. 12.

§ (7) DIFFRACTION.—A further consideration, limiting the use of geometrical optical theory, is the bearing of diffraction[1] upon the conclusions. A parallel beam of light emerging from a star to which the axis of a telescope is directed, is not brought to a point focus upon the axis, even if spherical aberration is corrected for all colours. It is distributed in a certain ring pattern over the whole focal plane, the central fact being that the diameters of the rings depend upon the ratio of the wave-length of the light to the aperture of the telescope. Thus F and B, denoting as before the focal length and semi-aperture of the telescope, λ the wave-length, and r distance from the axis, so that r/F is the circular measure of the angular radius of the ring in question, it is found that if we put

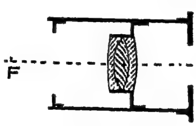


FIG. 13.

$$\frac{r}{F}=\left(\frac{\lambda}{2B}\right)\cdot\omega,$$

[1] See "Diffraction Gratings, Theory," § (4).

where ω is a numerical factor, we have zeroes of illumination for the values

$$\omega = 1{\cdot}220,\ 2{\cdot}233,\ 3{\cdot}238,\ 4{\cdot}241,\ 5{\cdot}243,\ \ldots$$

and within these rings respectively is comprised a fraction of the total illumination amounting to

$$\cdot 839,\ \cdot 071,\ \cdot 028,\ \cdot 015,\ \cdot 009,\ \ldots$$

while the maximum brightness occurs for the values

$$\omega = \cdot 000,\ 1{\cdot}635,\ 2{\cdot}679,\ 3{\cdot}699,\ 4{\cdot}710,\ 5{,}717,\ \ldots$$

and at these points the relative intensities are

$$1{\cdot}0000,\ \cdot 0175,\ \cdot 0042,\ \cdot 0016,\ \cdot 0008,\ \cdot 0004,\ \ldots$$

The only telescopic feature permitting control of this phenomenon is the aperture. If this is increased the angular diameters of the rings are all diminished in proportion. It is of interest to note that if the objective is reduced to an annulus by a central stop, the rings are somewhat diminished in diameter, but at the expense of much less favourable distribution of the proportions of the light between them, the shares of the outer rings being increased. Numerically, considering, say, a 10-inch object glass and the D-ray, the diameter of the spurious disc of a star lying within the first dark ring would be 0″·98. Hence the images of the members of a close pair would be more or less superposed if their distance was less than 1″; but much within this a good eye would detect the elongation of the united image, as indeed was found in practice by S. W. Burnham, who discovered when using a 6-inch objective by Alvan Clark many double stars the separation of which lay beyond this theoretical "resolving power."

Owing to the shorter wave-length of actinic rays it might seem that photography was better circumstanced than the eye for receiving small images, but it is well known that this is not the case. From a variety of reasons, the photographic image spreads; and indeed the diameters can be used as measures of the star's magnitude. If the smallest recorded images are of less diameter than 2″ or 3″, the results would as a rule be considered very favourable.

§ (8) Optical Glass.—So far we have spoken of the geometrical modification of the beam of light; we shall now consider briefly the characteristics of the glasses[1] out of which the lenses are constructed. Apart from the limitations which technical difficulties of glass manufacture impose, geometrical discussions are merely mathematical exercises. The earliest telescopes were made with a single lens as objective, and in consequence suffered severely from dispersion, the images for different colours being ranged along the axis. The focal length was made as great as possible, and, for example, D. G. Cassini discovered two of the satellites of Saturn, Tethys and Dione, with telescopes of 100 ft. and 136 ft. in length respectively. But since the separation as well

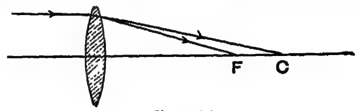


Fig. 14.

as the size of the images corresponding to, say, the rays C, F (*Fig.* 14) would increase proportionately with the focal length, it is an error to suppose that increase of focal length was any cure for chromatic faults, except in so far as the narrow beam does not allow spreading of any description. Achromatism was discovered and the principles published by John Dollond about 1760,[2] though achromatic lenses had been made before him by Chester Moor Hall, a private gentleman, as early as 1733. Making the converging lens of crown glass, the principle was to annul wholly the dispersion while only partly annulling the deviation by associating with the crown lens a divergent lens of greater proportionate dispersive power. This was found in "flint" glass or "English crystal," a dense and brilliant glass, which contained as base a silicate of lead. To obtain large homogeneous pieces of flint glass, free from striae, for negative lenses, proved a very difficult technical problem. It was solved by P. L. Guinand of Neuchatel, about 1800, whose secret was to stir the pot of melted glass with a fireclay rod to the last possible moment, and, allowing it to cool by itself, to take its natural fractures as marking off lumps of homogeneous constitution. When a large one was found this was softened again and moulded into the form of a disc. This secret he taught to Fraunhofer's firm in Bavaria and to Fraunhofer's successors, Merz and Mahler, by whom it was jealously guarded, so that in the middle of the nineteenth century they were the only people who could produce an objective of even so moderate a size as 8 inches diameter. Guinand, however, returned to Neuchatel, and his method passed to his son H. Guinand, and in succession to Feil, Mantois, and Parra in Paris, and through one of H. Guinand's collaborators, G. Bontemps, who took refuge in this country in the political disturbances of 1848, to Messrs. Chance in Birmingham. All the greatest lenses in the world, so far produced, excepting the 32-inch at Potsdam,[3] came from one or other of these firms. Their productions were, however, conservative, and a separate revolution in glass-making came from the researches

[1] See also article "Glass."

[2] *Phil. Trans.*, 1758, p. 733.

[3] Even these glasses, by Schott & Co., are of the old types.

of Abbe, Zeiss and Schott in collaboration at Jena, about 1880, investigating the effects of different constituents in glasses, with a view to finding stricter proportionality of dispersion over the whole spectrum, and generally to providing the greatest possible range and variety of refractive index and dispersion. In consequence of this the old terms *flint* and *crown*, though still employed, no longer imply the composition of the glass, nor its density, nor mean refractive index. The former term is reserved for glasses of high dispersion and the latter for glasses of low dispersion.

In spite of great technical improvements thus recently effected, defects in homogeneity of optical glass intrude themselves on the optician's calculations. Turrière states [1] that local anomalies of two or three units in the third decimal place are often found to affect refractive indices, whereas the calculations have been made assuming the constancy of the indices up to the fifth place. Moreover, it has not been found possible to avoid leaving some trace of mechanical strain from the process of annealing. This fault, which may be detected with the polariscope, will impair the definition with some traces of double refraction, which presumably would alter with the temperature.

We may take as representative of the glasses, out of which the great majority, at any rate, of the existing large refractors are constructed, Hard Crown (No. 605 of Chance's list) and Dense Flint (No. 361); the specifications of these glasses are the following, given as adhered to with considerable accuracy from one melting to another:

	Hard Crown.	Dense Flint.
Refr. index n_C	1·5150	1·6165
n_D	1·5175	1·6214
n_F	1·5235	1·6337
n_G	1·5284	1·6442
Dispersive power $(n_D-1)/(n_G-n_F)$	60·5	36·1
$(n_C-n_D)/(n_G-n_F)$	·294	·285
$(n_D-n_F)/(n_G-n_F)$	·706	·715
$(n_F-n_G)/(n_G-n_F)$	·554	·608

Glasses have been discovered in which the proportionality of dispersion is closer than for this pair, indeed practically complete, but they have proved liable to tarnish with exposure and age.

§ (9) Achromatism. (i.) *For Visual Use.*—In making the colour corrections there is one element in the construction which is disposable, the distribution of power between the crown and flint lenses, or two, if we add a small permissible separation of the lenses. The actual results attained are investigated by a method devised by Vogel, who set a direct vision spectroscope of small dispersion to receive the cone of rays from a bright star, say Vega or Arcturus. If the rays were brought to a focus the spectrum seen would be no broader than the star disc, and in fact for any setting of the spectroscope along the axis constrictions are seen which correspond to the wave-lengths in focus at that setting. The results—which agree precisely with what could be forecasted theoretically, as shown by H. D. Taylor [2]—are exhibited in the following table from measures of the Vienna 27-in. (Grubb, glass by Chance), the Pulkowa 30-in. (Alvan Clark, glass by Feil), and the Lick 36-in. (Alvan Clark, glass by Feil). The Vienna and Lick telescopes may be said to be identical, while the Pulkowa instrument can claim some superiority, but even for this it is to be noted that the image formed by the F-ray will be spread over a spurious disc 5″ in diameter, when the D light is brought to a point focus. Fine detail of varying colour, for example planetary surfaces, cannot be brought to a single satisfactory focus, and would suffer in definition accordingly.

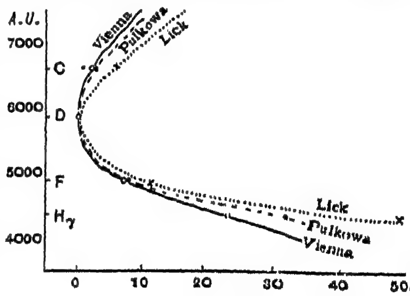


Fig. 15.

In the table on following page, *df* measures the setting required to bring into focus the line referred to, and *hff* is the angular diameter of the circle in which the cone of rays for that wave-length would meet the focal plane for the D-ray. See also *Fig.* 15.

The comment cannot be avoided that anything approaching true achromatism is impossible of attainment with the traditional objective of two lenses failing the production of types of glass that are quite unknown at present. The greater the telescope, with its increased light-grasp and wider separation of the foci, the more prominent the fault becomes. The optician has not enough scope allowed to produce his effects. As already remarked, H. D. Taylor has shown that a perfectly achromatic object glass can be made with three lenses, and has made such objectives up to any size for which the discs of glass were available. See *Fig.* 9.

[1] *Optique industrielle*, 1920, p. 25, a most useful work.

[2] *Mon. Not.* lv. 67.

COLOUR CURVES OF GREAT OBJECTIVES *

	Vienna.		Pulkowa.		Lick.	
Aperture	26·6 in. (67·5 cm.)		30·0 in. (76·2 cm.)		36·0 in. (91·4 cm.)	
Focal length . . .	408 in. (1036 cm.)		539 in. (1368 cm.)		674 in. (1712 cm.)	
Line.	*df.*	*h/f.*	*df.*	*h/f.*	*df.*	*h/f.*
	mm.		mm.		mm.	
C 6563 A.U. . . .	2·7	·0000170	3·0	·0000115	5·3	·0000162
D 5893 A.U. . . .	0·0	·0000000	0·0	·0000000	0·0	·0000000
F 4862 A.U. . . .	6·0	·0000378	6·4	·0000245	11·4	·0000349
H_γ 4341 A.U. . . .	23·5	·0001478	32·9	·0001258	48·3	·0001471

* Cf. Keeler, *Ast. Soc. Pacific*, ii. 258.

The lenses are made of three of Messrs. Schott's glasses, L_1 of Baryta Light Flint (O. 543, $n_D = 1·564$), L_3 of Light Silicate Crown (O. 374, $n_D = 1·511$), and L_2 of a Boro-Silicate Flint (O. 658, $n_D = 1·547$). With this arrangement, very simple compared to photographic combinations for camera lenses, faults of colour can be [1] and are completely corrected, and at the same time a field given which is free from coma and spherical aberration. The separation of the second and third lenses assists the correction of the chromatic difference of spherical aberration (Gauss's condition).

(ii.) *Photographic Lenses.*—The correction of a lens for photographic purposes is a simpler problem than one designed for visual use, since for ordinary plates the effective rays are confined to a much narrower region of the spectrum. The problem is otherwise similar. Of more complicated constructions the only one that has been made with the large apertures peculiar to astronomical work is the Taylor lens referred to above, p. 847, made of three separated units [2] (see *Fig.* 10, p. 848). The original of this lens has the following description : [3]

Aperture, 6·5 inches; focal length, 43 inches.

L_1	R_0 +10·64			
		t_1 = 0·83		n_D = 1·5180
	R_2 −72·45			
		t_3 = 4·39		
L_2	R_4 14·54			
		t_5 = 0·33		n_D = 1·6035
	R_6 +10·35			
		t_7 6·85		
L_3	R_8 +67·35			
		t_9 0·49		n_D 1·5180
	R_{10} −13·00			

The refractive indices are those of a silicate crown for the first and third lens, and a light flint for the middle lens. But the design possesses great flexibility. It has been made up to apertures of 10 inches, and to focal lengths as small as twice the aperture; further, it has given rise to an extensive series of derived forms. The lens is intended for photography, and visual distinctness is not demanded from the design.

§ (10) REFLECTING TELESCOPES.—The reflecting telescope claims three great advantages over the refractor: it deviates alike rays of every wave-length; its performance depends upon its surface only and not upon the interior constitution of its glass, nor on the chromatic balance of two separate glasses, nor diminishes in effectiveness as the scale increases owing to absorption; and the objective consists of only one surface, the correctness of which is assured by an easy and very sensitive laboratory test. Accordingly, as might be expected, in respect to dimensions and aperture ratio, reflectors have completely outrun refractors. While the greatest refractor is of 40 inches aperture with a ratio of aperture to focal length of 1 : 19, and is not likely at present to be superseded, there are in use reflectors of 60 inches (Mount Wilson), 73 inches (Victoria), and 100 inches (Mount Wilson), in each case with an aperture ratio of 1 : 5, and with accessory appliances which permit the same mirror to be used when required at enormous effective focal lengths of 100 feet, 134 feet, and even higher. If greater ones are desired there is no reason to suppose we have reached the limit in gain of power, or in possibility of construction, though the difficulties of temperature control become very formidable in these large sizes. The question then presents itself, why the refractor holds its own so well as it does. A few sentences will be devoted to this question later on.

Speculum metal ("Turner's Metal" copper, 4 atoms = 68·2 per cent, to tin, 1 atom = 31·8 per cent, as used by the Earl of Rosse, Rowland, and Brashear) has virtually gone out of use except for small pieces, because once it is tarnished the mirror must be remade. It is besides a worse reflector than silver. And it is heavy, and difficult to make and work. On the other hand, it is remarkably durable.

[1] For the calculation see H. D. Taylor, *Mon. Not.* liv. 328.
[2] H. D. Taylor, *Mon. Not* lxiv. 614.
[3] H. D. Taylor, *Applied Optics*, 1906, p. 182.

Some old mirrors, dating certainly from Herschel's day, are still in perfect order. But silver deposited on glass is the standard construction. The problem of figuring the glass correctly is simplified by the fact that there are no elements of construction to dispose of at will. In order to correct spherical aberration at the principal focus, the figure must be a paraboloid; distortion is absent, but coma cannot be removed, and astigmatism and curvature of the field are present as in the refractor.

In regard to the two aberrational features that are necessarily present both in the mirror and in usual objective doublet, we see from the particulars given above, § (3), that for the same aperture and focal length the radius of the focal circle, or the separation of the focal lines, is the same for both, while the radius of curvature of the field is equal to F for the mirror, while it is about three-eighths of this amount for the lens. As regards the coefficient of coma, it is no more than occurs in the Fraunhofer lens referred to in § (5) (iii.) ($\delta_2 G = -0{\cdot}500/F$, compared to $\delta_2 G = -0{\cdot}467/F$). An impression of a limited field of good definition attaches usually to reflectors, but this is derived solely from the large aperture ratio which is allowed to them.

The reflector is used in three forms: first, for direct photography without any second mirror, the plate holder being placed at the principal focus (*Fig.* 16); second, for visual and general work in the Newtonian form with a flat mirror placed at 45° to the beam, throwing the principal focus out to one side at the upper end of the tube (*Fig.* 17); and third, when very large scale and long focus are required, as a Cassegrain, with a convex mirror placed on the axis with its principal focus beyond the principal focus of the great mirror, so as to reduce the convergence of the beam. The great mirror is usually pierced with a central hole and the beam received on a plate holder or spectroscope behind it (*Fig.* 18).

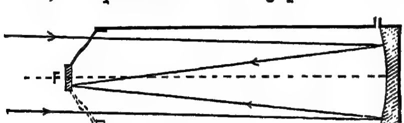


FIG. 16.

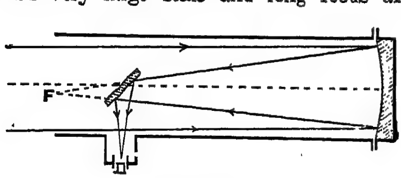


FIG. 17.

If F_1, F_2 are the foci of the two mirrors A_1, A_2, and F the emergent focus, the focal length of the combination is increased by the factor A_2F/A_2F_1. An ordinary value of this magnification might be fourfold. With this value the coma of the combination would be no more than belongs to a simple mirror of the same focal length. The curvature of the field would be increased, being about four times as great as for the original mirror: to correct for spherical aberration, if the mirror A_1 is parabolic, the section of A_2 requires to be an hyperbola, its measure e being about 3 where $e = 0$ gives the circle and $e = 1$ the parabola.[1]

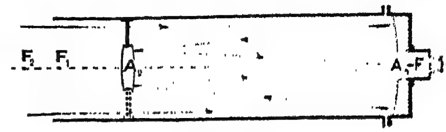


FIG. 18.

§ (11) TESTING A MIRROR. The possibility of producing truly paraboloidal or other figures in mirrors of large size depends upon having a sensitive and easy means of measuring a fault. This is found in Foucault's knife edge method (*Fig.* 19).

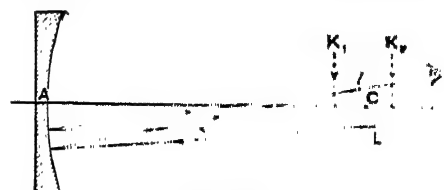


FIG. 19.

If C is the centre of curvature for A, L a point source or "artificial star" close to C, l the point of the beam reflected from any spot of the mirror, received by the eye at E beyond l, then if the beam is cut off gradually by a knife edge moving across from left to right at K_1 within the focus l, the spot of light on the mirror will seem to the eye to be obscured from right to left, but if it moves at K_2 the movement of shadow will be from left to right. Hence the position of the actual focus for any spot of the mirror can be determined precisely. A very good idea of the figure and faults of the whole mirror can be gained at a single glance, and if a series of diaphragms are prepared exposing the surface in successive zones, with very little preparation and with surprising certainty the centres of curvatures of the different zones can be laid down.[2] If the mirror is parabolic, these will not give the same point C, but will move towards the surface from C by an amount $\frac{1}{4}B^2/R$, supposing the source L is kept fixed during the observations.

[1] For a discussion of the correction of the Cassegrain see Sampson, *Phil. Trans.* ccxiii. 27.

[2] For an excellent account of the detail of this method see C. D. P. Davies, *Mon. Not.* lxix. 355.

A convincing test of the delicacy of the method is furnished if an assistant places his hand in the way of the beam near the mirror, when the undulations caused by the ascending currents of air its warmth produces will be seen immediately.

Moreover, the method is not restricted to visual application. Photographic records of the appearances have been obtained by J. Hartmann for the Potsdam 32-inch refractor,[1] and by P. Fox for the 40-inch Yerkes telescope,[1] which show upon examination much minute detail that would escape the eye; as, for example, epicycloidal traces of the motion of the polishing tool. These "focograms" have besides the invaluable quality that they can be compared with similar records of other instruments taken at other times.

A convex mirror, or a flat, or one of very small concavity, cannot be tested directly by this method. But it is usually possible to devise a combination with some converging system, the character of which is already proved, so as to test their performance. Or a flat may be tested with another known flat, by examining the formation of Newton's rings between them.

§ (12) Loss of Light in a Telescope.—We must now consider the whole loss of light, and the modification of its character in passing through a telescope.

(i.) *Refracting Telescopes.*—In transmitting light through glass surfaces, there is a loss at each surface by reflection. For direct incidence, theory indicates that the proportion so lost is $(\mu-1)^2/(\mu+1)^2$, where μ is the refractive index for the ray concerned; and for freshly polished glass this accords closely with experiment.[2] It may be taken, therefore, as between ·04 and ·05 per surface. In the case of objectives with separated lenses there will be four such occasions for loss; for small objectives the inner cemented surfaces do not give rise to any sensible loss, and therefore transmit some 10 per cent more of the light. The losses by absorption have been investigated by H. C. Vogel[3] for a number of glasses, including ordinary Silicate Crown and ordinary Light Flint (O. 203 and O. 340 of Schott's list), which are suitable for large telescopes. For the visual spectrum these glasses are very transparent, but O. 340 with a thickness of 10 cm. to 15 cm. transmits no light at all of wave-length less than 3760 A.U., and shows a sharply defined absorption band at 4186 A.U., and a weaker and diffuse one about 4370 A.U. The crown glass O. 203 also shows the former band. Otherwise the progression of absorption was regular and not strikingly different for the two glasses. Vogel gives a table from which the following is an extract, showing the amount of light transmitted by objectives of different thicknesses. In calculating for reflection, four surfaces of separation are allowed for. The aperture would, as a rule, be about six times the total thickness.

Thickness of Objective in cm.	Intensity of Transmitted in terms of Incident Light.			
	With allowance for Absorption only.		With allowance for Absorption and Reflection.	
	Visual Rays.	Actinic Rays.	Visual Rays.	Actinic Rays.
4	0·93	0·84	0·77	0·69
6	·90	·77	·75	·63
8	·87	·71	·72	·58
10	·84	·65	·70	·53
12	·82	·60	·67	·49
14	·79	·55	·65	·45
16	·76	·50	·63	·41
18	·74	·46	·61	·38
20	·71	·43	·59	·35

It will be noticed that the balance that may exist between visual and actinic rays in different sources of light will be transformed, and in different manners, by different objectives, so as to give, for example, inconsistent results in such research as that of the effective wave-lengths of stellar images.

(ii.) *Reflectors.*—Measurements of the relative intensity of light of different wave-lengths are acquiring great and growing importance, and it should be noted that reflectors of all types also exercise selective treatment of different rays. The following table (Hagen and Rubens) shows the percentage of light reflected from silver and a few other metallic surfaces. It will be noticed that silver deposited on the front of a glass plate is distinctly more effective than when deposited on the back, apart from absorption by the glass.

Percentage of Light reflected from Metals

Wave-length.	Silver.	Platinum.	Gold.	Speculum Metal.	Silver behind Glass.	Mercury behind Glass.
A.U.						
2510	34·1	33·8	39	30	..	..
2880	21·2	38·8	34	38	..	..
3050	9·1	39·8	32	42	..	..
3160	4·2	..	..	..	..	..
3260	14·6	41·4	28	..	..	..
3380	55·5	..	..	..	..	..
3570	74·6	43·4	28	51	..	..
3850	81·4	45·4	27	53	..	..
4200	86·6	51·8	29	56	..	..
4500	90·5	54·7	33	60	82	73
5000	91·3	58·4	47	63	84	71
5500	92·7	61·1	74	64	85	71
6000	92·6	64·2	84	64	85	70
6500	93·5	66·3	89	66	86	71
7000	94·6	69	92	67	87	73
8000	96·3	70·3	95	71	..	..
1000	96·6	75·5	97	74	..	..

[1] *Astrophysical Journal*, xxvii. 258 and 248.
[2] Cf. Rayleigh, *Collected Works*, ii. 522.
[3] *Astrophys. Journ.*, 1897, v. 75.

A silver surface, therefore, cannot be approached by any other, so long as it remains in good order, both for the visual and the actinic region of the spectrum. But in the ultra-violet, about λ3160, for a short range, it becomes almost transparent, the percentage of light reflected being hardly more than from clear glass. This limits its application, for example, in some photo-electric researches. But in the same region an ordinary large object glass is completely opaque.

§ (13) General Considerations.—It will be realised that some of the defects of a telescope considered in the foregoing pages are irremovable with the means allowed, and some of the requirements are mutually incompatible. But usually in each observation one or another consideration will be predominant.

So long as coma is not aggressive, a wide field may be dispensed with in a large number of observations: for example, in transit work, where a good field of 15′ radius gives ample time to take the observation, or in double star measures and micrometer work generally; and in all other classes of direct visual work where the object is brought to the centre of the field for examination; and in work with a slit-spectroscope, whether visual or photographic. Equally in these cases curvature of the field and astigmatism play no part.

In some classes of photometric work no good image at all is wanted; the star is deliberately thrown out of focus, preparatory to estimating its total light.

Again, photographic work, where a correct delineation of a field of stars is required, may be divided into different classes according to the scale or focal length. With a focal length of 20 feet, a field 1° across covers a circle 5 inches in diameter. It would be easy to secure a correct field of so small a diameter, and more would not be asked of such an instrument, if only because the plates would be inconveniently large. If we want large fields we go to the other extreme and employ, as, for example, for the Franklin-Adams chart, a Cooke lens of aperture 10 inches and focal length 45 inches, which gives sensibly correct pictures of 15° diameter, with a flat field. There would be some distortion in such a picture, and the scale would be comparatively small. But the plate is not used for measurement but for enumeration; or it is used for photography of nebulae or comets, where increase of scale without increase of aperture loses the detail by diffusing it.

Midway between these extremes we have the telescopes of the Astrographic Catalogue; these are each, as near as may be, of 3438 mm. focal length, so that on the image 1 mm. represents 1 minute of arc; a field 2° square is photographed, which allowing for margin gives a plate about 6 in. × 6 in. On this scale over ten thousand plates are required to cover the sky—the mere difficulties of numbers will speak for themselves. The object glasses, 340 mm. = 13·5 in. diameter, were specified to be free from coma, but astigmatism and curvature of the extreme images are easily marked, the former by the images taking the form of two short crossed lines, and the latter by the lower density of stars recorded, owing to fainter stars failing to impress themselves when their light is a little diffused. Such a plate can be measured right up to its limits for the relative co-ordinates of the stars, but is quite unsuited for work of the highest accuracy, for instance for stellar parallax.

The reflector is well suited for all classes of visual observation, except transit work; and particularly suited for the spectroscope, owing to its light-gathering power. It is unapproachable for photographs of nebulae, but has been but little used for fields of stars designed for measurement, owing to the irremovable coma which vitiates its field, and which becomes prominent in the large aperture ratios with which the reflector is usually made. In somewhat the same way methods and instruments must be selected so as to avoid the intrusion of residual chromatic faults. A great refractor is usually applied to faint objects; if applied to bright ones, there is a blue glare in the field of disturbing intensity. This may be removed by using for such observations a colour screen which shuts out the blue rays. In the same way, by excluding rays that are not wanted, very perfect photographs have been obtained with the Yerkes 40-in. refractor, which is corrected visually, by interposing a yellow colour screen and using yellow sensitive plates. The loss of speed is of course very great. Again, a telescope corrected visually can be applied to photography, or conversely, by interposing a lens in the beam which collects to a focus the rejected rays. All the field except the centre is thereby sacrificed, but this does not matter if the purpose is to employ a slit-spectroscope.

As remarked above, a region of investigation which promises to occupy growing importance is the distribution of light in different parts of the spectrum. This is a matter that has hardly been considered, in view of its importance. All telescopes in use modify the distribution they record, each in a different way, and reflectors in a way that changes with the condition of the silvered surface.

§ (14) Mounting of Telescopes: The Transit Instrument.—We shall now describe briefly the principles of mounting telescopes, but, as explained in the beginning of this article, we shall avoid as far as possible astronomical detail, and direct attention to the geometrical and mechanical problem.

The Transit Instrument is designed for measuring relative position of points on the

sky, or, given these, the orientation of the observer in latitude and hour angle with respect to them. The two problems are logically interlocked, but by a process of sifting, familiar in astronomy, they are gradually and effectively separated from one another. The necessity for the measuring instrument is the greatest fixity, subject to commanding the whole sky. The ordinary standard form aims at sweeping the meridian with the optic axis. This requires (1) two pillars built up from the ground, carrying "Y"s at the same level and standing E. and W. with respect to one another; (2) a massive axis, with its ends resting on the "Y"s as bearings; (3) a telescope fastened at right angles to this axis; this telescope is made in two symmetrical pieces, the object glass and eye end being interchangeable if desired, and bolted to the "cube" which forms the middle of the axis, and which is perforated to allow the beam of light to travel through. General stability and constancy is sought by making all the supports massive, and relieving by counterpoises on adjustable springs the greatest part of the pressure on the "Y"s. There are, however, temperature or other seasonal changes to allow for. The former are dealt with as far as possible by considering symmetry as a rigorous requirement of the design. In the Cape Transit Circle as well, to guard against local differences, the telescope proper is completely enclosed in a copper sheath, and the instrument house moves bodily away from it in two halves. But finally it is assumed in the resulting observation that the telescope shows traces of complete maladjustment, *i.e.* three degrees of freedom, the coefficients of which can only be derived by special check observations. Particulars of the processes would carry us too far into astronomical details. The print (*Fig.* 20), derived from Messrs. Troughton & Simms's Catalogue, shows the Cape Transit Circle (6-in. objective), and examination of it will show how the principles are put into practice.

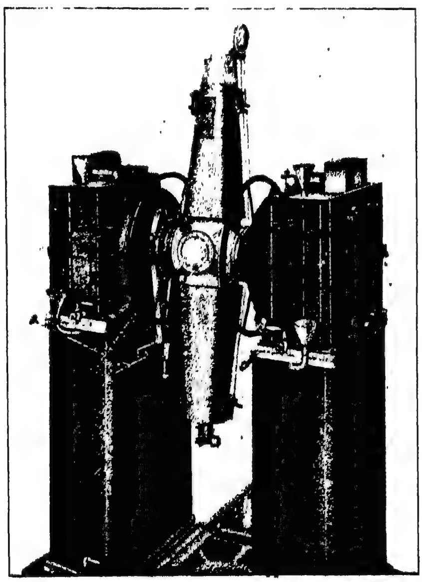


FIG. 20.

§ (15) EQUATORIAL.—For the examination of an individual object, or for photographing or measuring a particular field, a certain time up to an hour or several hours is required, and the telescope must be endowed with motion which cancels the earth's rotation. That is to say, it must rotate east to west about an axis parallel to the earth's poles, at a rate which would give one revolution per day. The construction, then, is that of a massive axis pivoted so as to lie parallel to the earth's axis and geared by a clamp when desired to clockwork which gives it the necessary motion. This polar axis carries perpendicular to itself the declination axis, about which the telescope itself is capable of rotating. If the telescope is turned about the declination axis, and the polar axis turned until a desired object is in the field of view, and then clamped, the clockwork will keep it pointed on the object. There is a variety of ways in which these purposes may be carried out.

(1) The polar axis may be prolonged beyond its bearings and terminate in a pair of journals in which the declination axis swings; (2) the polar axis may be pivoted at its extremities and be built up in the form of a frame or cage in the midst of which the telescope swings upon its declination axis; (3) the declination axis may be attached in two separate pieces to opposite sides of the telescope and, broken itself, support the telescope symmetrically; (4) the declination axis may be prolonged on both sides of the polar axis, carrying the telescope attached by bands to a cradle upon one side, with a counterpoise upon the other. All these forms are made and are illustrated in *Figs.* 21, 22, and 23 on p. 856.

These illustrations show the constructions of the three great reflectors made in recent years. The weights of the moving parts of these instruments are very great. For the Canadian telescope (73") at Victoria, Vancouver Island, the polar axis weighs 10 tons, the declination axis 5, the mirror, cell, and tube together 12 tons, and the whole instrument 55 tons. To bring pieces of such size within the range and accuracy of "fine" mechanical work is a great feat of exact engineering. In the case of the

two Mount Wilson telescopes the weight is carried chiefly by large discs plunged in baths of mercury; the Victoria telescope relies upon ball bearings, and the counterpoising is so perfect and the movement so frictionless that a pressure of 3 lbs. weight applied at the end of

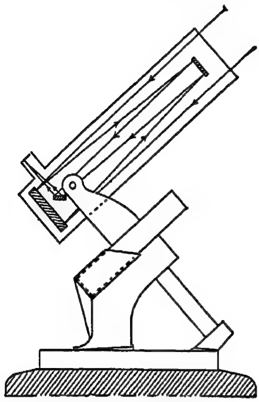


FIG. 21.—Fork, 60″.

the declination axis suffices to turn it. In these great telescopes, designed for grasping very faint objects, it is essential that the clockwork driving should be as perfect as the optical work, otherwise the image on the photographic plate is diffused and the advantage of size is lost. Driving differs from the optical

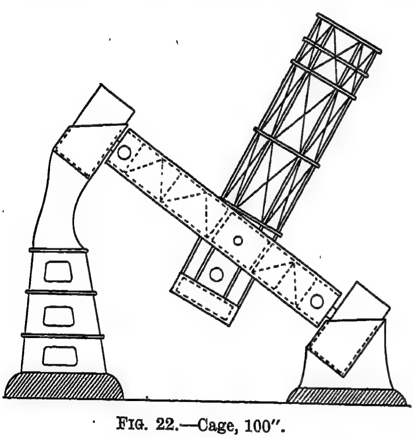


FIG. 22.—Cage, 100″.

work in this, that it is subject to control. A star is kept under visual observation, bisected by a pair of cross-wires; when a bright star is available it is a good plan to form a disc by throwing it out of focus and bringing this disc to bisection by two heavy wires which are visible without illuminating the field. This visual supervision is usually made with an accessory "guiding" telescope fixed to the same frame, but the most perfect way is to have the plate holder constructed with an eyepiece attached to it so that the actual set of the plate may be observed directly. When the clock drive requires to be accelerated or retarded, it is essential that this should be done with no jar whatever. Differential gear thrown in and out of action accomplishes it. If the clock carries an arbor making one revolution per second, the standard clock may be made to exercise automatic control in applying the

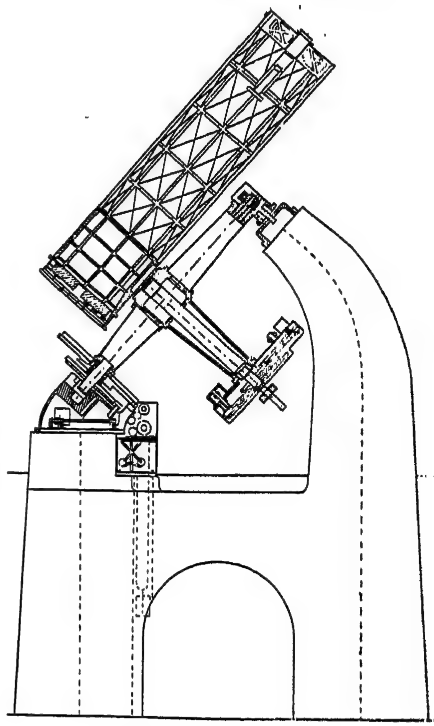


FIG. 23.—Counterpoise, 73″.

gear, though this does not dispense with personal watching. For automatic control, a simple gear, which is said to work well, is a sector of soft iron, attached on the one-second arbor, and revolving so as to pass near the poles of an electromagnet which a signal from the clock makes live once every second. But most of the methods of automatic control are prejudicial to smooth running.

The mounting of a refractor is nearly always made by prolonging the declination axis beyond its journals and attaching the declination axis as a cross-head, carrying the telescope at one end and a counterpoise at the other. For this form there are two, reversed, settings for any object, like the "changes of face" of a theodolite, and, generally speaking, one or

the other setting may be inaccessible. The cross-head should not be too short, otherwise in some positions the eye end will foul the supporting pillar and cut short an observation by necessitating change of face in the middle of it.

Fig. 24 shows a good example of modern construction, an 18-inch visual telescope for the observatory of Rio de Janeiro, by T. Cooke & Sons. Beside the great telescope are two cameras, each of 70 inches focal length and furnished with 10-inch Cooke lenses. For the use of these cameras a 10-inch follower is provided, or the great telescope itself could serve as a visual guider.

Consideration of these designs will show that the refractor permits a more symmetrical, compact, and stronger mounting than the reflector, the tube being closed at both ends. Its optical condition is also not liable to variation. Both these are very important considerations, and will account for a certain amount of just preference for the refractor. A third consideration is less reasonable but is probably present. Difficult as the optical problem of making the object glass of a telescope is, in comparison with the mechanical problem of mounting it, the most expensive part of a telescope arises in building its dome, mount, and clockwork. A mirror is very much easier and cheaper to make than an object glass of the same power, and when so much is spent upon accessories there is perhaps some inclination to complete them with the more costly optical provision. Generally, one would say that a reflector deserves to have an even greater proportion of use in large optical work than it gets at present.

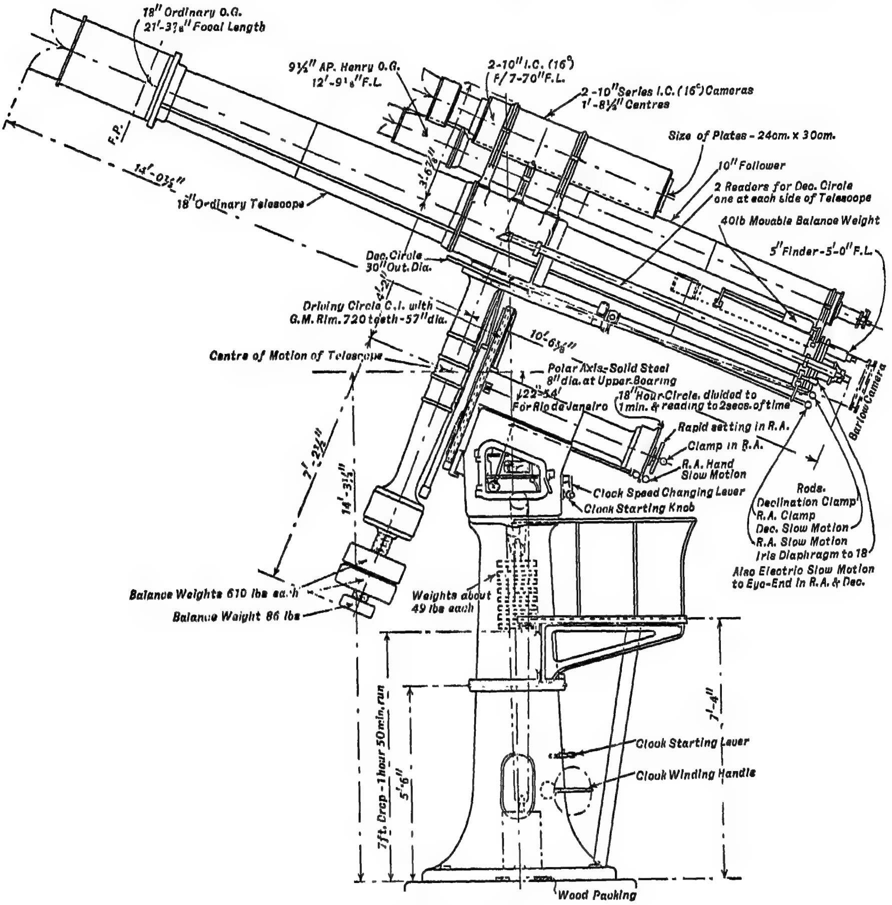


FIG. 24.—Rio, 18″.

§ (16) THE COELOSTAT.—In place c
ing the daily rotation with the
there is an alternative of keeping the
fixed and reflecting the rays into it

sometimes followed in the case of the sun, where there is ample light, because the use of long focus lenses and spectroscopes presents fewer difficulties if they are kept in a horizontal (or a vertical) plane. It is accomplished by one of two methods. The simplest device is the coelostat (*Fig.* 25). If S is any point upon the sky, and N gives the direction of the normal to a plane mirror, then if SNS′ is a great circle and SN = NS′, S′ gives the direction of the reflected ray which arrives from S. Hence it is plain that if N is made to move along the equator with half the speed of the daily rotation, the ray S may be reflected to any fixed point S′ lying on the small circle of south polar distance equal to the north polar distance of S; in particular it may be sent horizontally to one or other of the two points S_0' in which this circle cuts the horizon. And the image of the field round S will remain stationary and not rotate as S moves across the sky. On the other hand, the horizontal direction in which the ray is sent is not open to choice. To send it in a prescribed direction a second, fixed, plane mirror would be required.

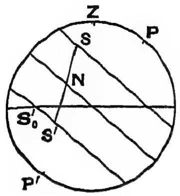


FIG. 25.

§ (17) THE HELIOSTAT.—Foucault's Heliostat, on the other hand, will send the beam constantly in any chosen direction (*Fig.* 26). AB is directed to the pole, TN is the normal to the mirror; then if TB = BN, a ray arriving in the direction BN is reflected from the mirror in the direction BT. Hence if there is a universal joint at T, and the rod TN slides through a well-fitting sleeve at N, and BN is clamped to AB so as to point at the polar distance of the sun, then giving the united piece AB + BN a movement equal to the apparent daily rotation, the mirror will send the beam out in the direction BT, which may be any desired direction. The universal joint at T is made by furnishing the cell of the mirror with a pair of horizontal pivots which rest on the two arms of a large solid fork, which itself is capable of free rotation about a vertical axis. A disadvantage attending the Heliostat is that the image of the sun which is formed rotates, the north polar point of the sun passing from one side of the vertical to the other.

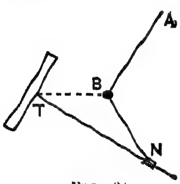


FIG. 26.

R. A. S.

TELESCOPE OBJECTIVES, TESTING OF. See "Objectives, Testing of Compound."

TELESCOPES, SMALL

§ (1) GENERAL PROPERTIES. Telescopes are optical instruments designed to enable comparatively distant objects to be observed on a uniform magnified scale. The image formed on the retina is thus increased in a ratio independent of the distance of the object, and the value of this ratio is called the magnifying power of the instrument. Paradoxical as it may seem, this result is achieved by presenting for inspection to the eye a diminished virtual image of the object, the ratio of the size of this image to that of the object being the reciprocal of the magnifying power. The explanation of this paradox whereby the image on the retina varies inversely as the size of the immediate object viewed directly by the eye is vital to a proper understanding of the telescope.

It follows from the ordinary theory of refraction through a lens[1] that if x is the separation parallel to the axis of two object planes which have images magnified in the ratios G_1 and G_2 respectively, and if x' is the corresponding separation between the images,

$$x' = x \, . \, G_1 \, . \, G_2,$$

provided the refractive indices of the object and image spaces are equal. If, now, the case of a telescope of magnifying power P is considered, the image of every transverse object is diminished P times, so that $G_1 = G_2 = 1/P$, and if y is the length of a transverse object the image is of length y' where $y' = y/P$. Among the objects in the object space is the first lens of the telescope, and in nearly all cases the eye is placed as nearly as possible at the image of this lens. The distance of any particular object from this lens may be identified with x in the above formula, and x' is then the distance of the image from the eye. The relation between them is

$$x' = \frac{x}{P^2},$$

that is to say, the apparent distance of the object from the eye has been diminished as the square of the reduction in the linear dimensions of the object. Thus, although the image is less than the object, it is so much nearer the eye that the final result is that the telescopic image appears to have been enlarged in the ratio in which, in fact, it has been diminished. Put in another way this is a particular illustration of the well known law that the angular magnification varies inversely as the linear magnification.

The relation between x and x' shows that the range of ocular accommodation required

[1] See "Lenses, Theory of Simple," § (7); also "Optical Calculations," § (7).

in viewing a perfect image in a telescope is equal to that called into play when the same objects are viewed by the eye from a nearer point, such that the retinal images are equal to those seen from the further view point through the telescope. The only difference, in fact, apart from any due to imperfections in the instrument, is the incorrect perspective in the telescopic view for an image of the apparent size obtained by the use of the telescope. Owing to instrumental defects a much greater range of accommodation is, in fact, required, and special provision is made for this purpose in the telescope focussing adjustment.

It will be gathered from what has already been said that the telescope is peculiar among optical instruments in forming images of all real objects at a finite distance from the instrument. Thus it is only when celestial bodies are under examination that the emergent beams of light are parallel with an instrument in normal adjustment. Notwithstanding this, it is permissible, in dealing with the construction of the telescope, to consider only the light emanating from infinitely distant objects. The image of an infinitely distant object being itself at infinity, the condition characteristic of telescopes in the notation used in "Optical Calculations" is evidently

$$K=0,$$

and equations (31) of that article then show that the magnification is independent of the distance of the object, and since the magnifying power is the reciprocal of the linear magnification, the desirable conditions for all rays in terms of the magnifying power Γ are

$$\frac{\partial K_{1,n}}{\partial K_1}=\Gamma, \quad \frac{\partial K_{1,n}}{\partial K_n}=\frac{1}{\Gamma}.$$

From equations (33) it is seen that the relations between the incident and emergent direction cosines of the ray are

$$\frac{M'}{M}=\frac{N'}{N}=\Gamma,$$

and not the tangent law which is usually assumed to hold, viz.

$$\frac{M'}{M}=\frac{N'}{N}=\frac{L'}{L}\Gamma.$$

A further quantity of importance is the magnitude of $\partial^2 K_{1,n}/\partial K_1 \partial K_n$, or D; this is of the dimensions of a length, and may be interpreted in terms of the distance of the centre of rotation of the eye from the last surface of the telescope when the eye is in its most favourable position. The value of this eye distance is D/Γ, and similarly the image of the last surface of the telescope in the rest of the system lies in the object space at a distance DΓ from the first lens. D may be given any finite value that is desired by changing the scale of the whole instrument, so it is necessary in considering its value to consider the length of the complete instrument, or alternatively the focal length of a selected component lens. The value of D in comparison with the apertures of the lenses determines the field of view of the telescope.

§ (2) Galilean Telescope.—The simplest telescope consists of two lenses, the essential relations being in terms of paraxial quantities

$$\kappa_1=\frac{\Gamma-1}{t\Gamma},$$

$$\kappa_2=\frac{1-\Gamma}{t},$$

$$D=-t,$$

so that

$$\frac{1}{\kappa_1}+\frac{1}{\kappa_2}=t.$$

If Γ is positive, so that the telescope presents erect images to the eye, one lens is positive and one negative, the image of either lens formed by the other being virtual. This is the construction used in the Galilean telescope (*Fig.* 1). The chief merits of the instrument, apart from

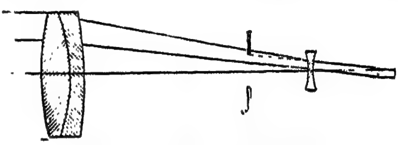


Fig. 1.—Galilean Telescope.

its cheapness, are the brightness of the field due to the small absorption of light. The chief defects are the small field of view resulting from the negative value of D. The chief use of this telescope is for low magnifications —say 2 to 5—and for night observation at sea. In an average instrument the "apparent field," measured by the angles corresponding to the extreme values of M′ when N′=0, is about 16°, the value in practice being nearly independent of the magnification. In theory it should be possible to obtain apparent fields which increase as the magnification falls. The real field, which corresponds to the extreme values of M with N=0, is evidently obtainable with sufficient accuracy by the relation

Apparent field = Real field × Magnifying power.

§ (3) Astronomical Telescope.—If the image presented to the eye is inverted, corresponding to a negative value of Γ, both lenses are positive, and the desirable position for the eye is external to the instrument. Considerable advantages, however, result in this case by utilising three separate lenses, as

described below. An ordinary form is shown in *Fig.* 2. It will be observed from the preceding formulae that

$$\kappa_2 = -\kappa_1 \Gamma,$$

and this result may be extended to telescopes generally, however complicated their structure.

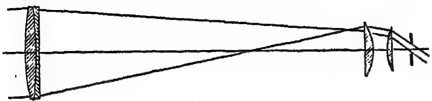


FIG. 2.

In fact, the equations already referred to give at once for all values of λ

$$\Gamma = -\frac{K_{\lambda+1,n}}{K_{1,\lambda}} = \frac{\partial K_{\lambda,n}}{\partial K_\lambda} \bigg/ \frac{\partial K_{1,\lambda}}{\partial K_\lambda}.$$

When there are three lenses the two separations, t_1 and t_2, and D may be given arbitrary values. The powers of the lenses are then given by

$$\kappa_1 = \frac{1}{t_1} + \frac{1}{D\Gamma} + \frac{t_2}{Dt_1},$$

$$\kappa_2 = \frac{1}{t_1} + \frac{D}{t_1 t_2} + \frac{1}{t_2},$$

$$\kappa_3 = \frac{1}{t_2} + \frac{\Gamma}{D} + \frac{t_1}{Dt_2}.$$

For comfortable use D should not be numerically less than 25Γ mm., its sign agreeing with that of Γ. If the apertures of the lenses are denoted by a_1, a_2, a_3 respectively, the maximum useful aperture under normal conditions is determined by

$$a_1 = e|\Gamma|,$$

where e is the largest diameter of the pupil of the observer's eye when using the instrument. Under ordinary conditions a value for e of about 6 mm. may be adopted. This determines the aperture of the objective, and a knowledge of the performance of objectives of various relative apertures enables an upper limit to be assigned to the value of κ_1. In good instruments values of $a_1\kappa_1$ as high as 0·3 or 0·35 may be found. The remaining quantities follow readily on considering the field of view. The useful rays are effectually limited by two stops in the object space, the images in the front part of the instrument of the second and third lens apertures. The image of the third lens, as has been seen, is of diameter $a_3|\Gamma|$, and is a distance $D\Gamma$ in front of the first lens. The image of the second lens is of diameter

$$\left|\frac{a_2 D\Gamma}{t_1 + t_2\Gamma}\right|,$$

and is situated a distance

$$\frac{t_1 D\Gamma}{t_1 + t_2\Gamma}$$

in front of the first lens. As $t_1 + t_2\Gamma$ tends to be small it is simpler to consider the latter limit as a cone through the first lens aperture of semi-angle $\tan^{-1}(a_2/2t_1)$. The effect of the two apertures is to enable the whole aperture of the objective to transmit useful light for an angular field determined by the smaller of the two ratios a_2/t_1 and $|a_3/D| - |a_1/\Gamma|$, with an outer zone of the field corresponding to the space between this limit and the smaller of the two ratios a_2/t_1 and $|a_3/D| + |a_1/\Gamma|$. In this zone the brightness of the image continuously diminishes from the inner to the outer limit because only a diminishing portion of the area of the objective receives light which is eventually transmitted to the image space. It is customary to include about one-half of this outer zone in the field of view, and a metal diaphragm is placed where the real image of distant objects at the limiting angle is formed to give a sharp edge to the field. The case of practical importance is when both D and Γ are negative and the powers of all the lenses are positive. A telescope of this kind is the foundation of the well-known prismatic telescopes usually combined in pairs for binocular vision. A positive value is always secured for $t_1 + t_2\Gamma$, and this has a direct influence on the correction of the component lenses for colour. The system will be automatically maintained in telescopic adjustment for another colour if t_1 and t_2 remain unchanged in the expressions for κ_1, κ_2, κ_3 while D and Γ are changed. Suppose that D and Γ become $D(1+d)$ and $\Gamma(1+\gamma)$ where d and γ are small. If the new value of each κ is obtained by introducing a factor $(1+1/\nu)$, evidently the first approximations are

$$\frac{\kappa_1}{\nu_1} = -\frac{d+\gamma}{D\Gamma} - \frac{t_2 d}{Dt_1},$$

$$\frac{\kappa_2}{\nu_2} = \frac{Dd}{t_1 t_2},$$

$$\frac{\kappa_3}{\nu_3} = -\frac{\Gamma(d-\gamma)}{D} - \frac{t_1 d}{Dt_2}.$$

Now small variations in D with the colour are not of serious importance, so that the second lens need not be corrected for colour. If this lens is of a crown glass the variation of D over the visible spectrum may be of the order of 4 mm. If the value of d in terms of ν_2 is substituted in the other equations, it is found that

$$\frac{\kappa_1}{\nu_1} = -\frac{\gamma}{D\Gamma} - \frac{(t_1 + t_2\Gamma)\kappa_2 t_2}{D^2\Gamma\nu_2},$$

$$\frac{\kappa_3}{\nu_3} = \frac{\Gamma\gamma}{D} - \frac{(t_1 + t_2\Gamma)\kappa_2 t_1}{D^2\nu_2}.$$

Obviously the image will be visibly imperfect unless γ is very small, and the proper principle

for the correction of the external lenses is shown by making it zero. Since Γ is negative and all the other terms positive, ν_1 should be positive and ν_3 negative, that is, the objective should be under-corrected for colour and the eye lens over-corrected. If the sign of $t_1+t_2\Gamma$ were reversed, so that the eyepiece were of the Huygenian rather than the Ramsden type, the requirements as regards these two lenses would be reversed also. The determination of the precise data for the construction of these lenses can follow the lines described under "Optical Calculations." When prisms are inserted for the erection of the image between the lenses the effect of the thickness of the glass—equivalent to a variation in t_1—must be taken into account. It is worth noting that when the second and third lenses which form the eyepiece are regarded as a separate instrument, the conditions for the chromatic correction which follow from the above equations when both these lenses are of the same glass are inconsistent with that usually given, viz.

$$2t_2=\frac{1}{\kappa_2}+\frac{1}{\kappa_3}.$$

The correct condition in this case is

$$2t_2=\frac{1}{\kappa_2}\left(1+\frac{t_2}{t_1}\right)+\frac{1}{\kappa_3},$$

and this approximates to the usual form when t_2 is small compared with t_1. The desirable separation on this theory is always greater than that given by the ordinary rule, and reaches a value twice as great when the first principal focus of the field lens is situated at the objective.

§ (4) Terrestrial Telescope.—Telescopes having four or more separated lenses may be discussed in a similar way. The principal aim in introducing additional lenses is usually to present an erect instead of an inverted image to the eye, so that a form with positive values for D and Γ is desired. It will suffice, by way of example, to consider a system of five separated lenses.

Let $$t_rt_{r-1}\kappa_r-t_r-t_{r-1}=k_r,$$

and regard k_2 and k_4 as independent variables. Then the powers of the odd lenses are determined from

$$\kappa_1=\frac{1}{t_1}+\frac{1}{D\Gamma}+\frac{t_2}{t_1k_2}+\frac{k_4}{Dk_2},$$

$$k_3=\frac{t_1t_3}{k_2}+\frac{t_2t_4}{k_4}+\frac{Dt_2t_3}{k_2k_4},$$

$$\kappa_5=\frac{1}{t_4}+\frac{\Gamma}{D}+\frac{t_3}{t_4k_4}+\frac{k_2}{Dk_4}.$$

In practice there are limitations to the extent to which k_2 and k_4 may be varied independently of one another, and a restriction it appears natural to adopt is that k_3 will be small, so that the even lenses, which are likely to be field lenses, will be approximately in conjugate planes with respect to the third lens. As D is to be positive this implies that k_2 and k_4 will differ in sign. A further limitation the designer will wish to bear in mind is the value of the sum of the powers of all the lenses. If this is too large the curvature and astigmatism will tend to be inconveniently great. The value of this sum is

$$\Sigma\kappa=\frac{(D-k_2-k_4)^2}{Dk_2k_4}+\frac{(1-\Gamma)^2}{D\Gamma}+\frac{t_1t_2\kappa_2^2}{k_2}+\frac{t_3t_4\kappa_4^2}{k_4},$$

two terms being positive and two negative. In special cases these expressions are unsuitable, for k_2 or k_4 may be zero. Ambiguity in these instances may be avoided by treating k_3 as an independent variable, and the formulae become

$$\kappa_1=\frac{1}{t_1}+\frac{1}{D\Gamma}+\frac{k_3k_4-t_2t_4}{Dt_1t_3},$$

$$\kappa_5=\frac{1}{t_4}+\frac{\Gamma}{D}+\frac{k_2k_3-t_1t_3}{Dt_2t_4},$$

$$\Sigma\kappa=\frac{(1+\Gamma)^2}{D\Gamma}+\frac{1}{Dt_1t_2t_3t_4}\{k_3(k_2t_1t_3+k_4t_2t_4)-(t_1t_3+t_2t_4)^2\}+\frac{k_2}{t_1t_2}+\frac{k_3}{t_2t_3}+\frac{k_4}{t_3t_4}+\frac{2}{t_1}+\frac{2}{t_2}+\frac{2}{t_3}+\frac{2}{t_4}.$$

If $k_2=0$,

$$k_4=-\frac{Dt_2}{t_1},$$

and similarly, if $k_4=0$,

$$k_2=-\frac{Dt_3}{t_4}.$$

To determine the character of the chromatic correction necessary in the components, it will be observed that if latitude is allowed in D control of the two external lenses alone is necessary to ensure that the system remains a telescope of magnifying power Γ as the colour changes. It may, in the first place, be assumed that the even lenses are of a single glass of the same melting. In the particular case

$$k_2=k_3=0,\quad \nu_2=\nu_4=\nu,$$

$$\Sigma\kappa=\frac{(1-\Gamma)^2}{D\Gamma}+\frac{2(t_1+t_2)(t_2+t_3)}{t_1t_2t_3}-\frac{(Dt_2-t_1t_3+t_2t_4)^2}{Dt_1t_2t_3t_4},$$

$$Dt_2t_3(1-\nu d)=(t_1+t_2)t_3t_4+(t_3+t_4)t_1t_2,$$

indicating a negative value for νd. Also

$$\left(\frac{1}{\nu_1}+d\right)\left(\frac{1}{t_1}+\frac{1}{D\Gamma}-\frac{t_2t_4}{Dt_1t_3}\right)=\frac{d}{t_1}-\frac{\gamma}{D\Gamma}-\frac{t_2(t_2+t_3)}{t_1^2t_3\nu_2},$$

and $$\left(\frac{1}{\nu_5}+d\right)\left(\frac{1}{t_4}+\frac{\Gamma}{D}-\frac{t_1t_3}{Dt_2t_4}\right)=\frac{d}{t_4}+\frac{\Gamma\gamma}{D}.$$

Thus if $\Gamma = t_1t_3/t_2t_4$, the fifth lens requires to be achromatic, and either the first and third must be separately achromatised or one must be under- and the other over-corrected, the relation between the amounts being

$$(t_2 + t_3)\nu_1 + \Gamma t_4 \nu_3 = 0.$$

It is preferable to depart somewhat from this condition, so avoiding having lenses in sharp focus with the image. If the real images precede the second and fourth lenses, the fifth lens must be over-corrected for colour but the first and third lenses must form an under-corrected combination. If the images follow the second and fourth lenses the fifth will be under-corrected, and the first and third together will be over-corrected. The modifications involved when k_3 is not zero are easily traced, and on the principles already illustrated lead readily to the justification of the form adopted for the terrestrial eyepiece when the last four lenses are of a single glass. When large fields of view are desired the third lens should be built from more than one glass to enable the spherical aberrations to be properly controlled.

Telescopes consisting of five separated lenses are largely used as gun-sights. In this case one of the real images lies in a plane in which cross lines are placed, either fine metal wires or lines etched on a flat glass surface. Provision is made for the illumination of the lines for night use, the light falling on the wires from the eyepiece side, or in the case of etched lines, entering the glass plate through a polished edge.

§ (5) Prismatic Telescopes and Binoculars.—The three-lens inverting telescope is usually converted into an erecting telescope by the insertion of erecting prisms. Many different forms have been employed, in some of which the prisms lie between the lenses, while in others they precede them. In a few military instruments the prisms are also utilised to change the direction of sight, as in gun-sights to be used against aircraft. In the most usual form (*Fig.* 3) two isosceles right-angled prisms are used, with their principal planes at right angles, the light entering and leaving each prism by the hypotenuse and being reflected at each of the equal faces. The effect of each prism is to reverse the direction in which the light travels and to invert the image in the principal plane of the prism, but not in the perpendicular plane. The two prisms together thus produce the inversion required while the light emerges travelling in its original direction, though the emergent axis is of course displaced relatively to the incident axis. Advantage is taken of this displacement in binocular telescopes to secure as large a separation as possible between the centres of the objectives while obtaining the eyepiece separation necessary for comfortable use with both eyes. The effect is to increase greatly the stereoscopic power of the instrument, that is, the extent to which the varying distances of objects from the observer is made evident by the slight eye adjustments necessary to bring their images on corresponding retinal areas. It is easily seen that, referred to the unaided eye as a standard in which this power is represented by unity, the stereoscopic power is equal to

$$\frac{\text{Magnifying power of telescope} \times \text{Separation of objective centres}}{\text{Separation of eyepiece centres}}.$$

The action of the prism is very easily determined geometrically in simple cases, and in complex examples a polar diagram enables the effects of any number of reflections to be traced without difficulty.

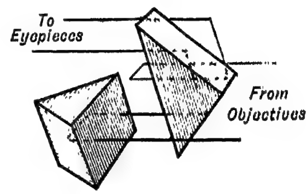


Fig. 3.

When there are two plane reflecting surfaces inclined to one another at an angle α, as in *Fig.* 4, the figure illustrates that any ray incident in a principal plane is deviated through an angle 2α. When the ray does not lie in a principal plane of the figure its projection on the principal plane follows the above law, and its component parallel to the reflecting planes is unaffected. If, then, α is a right angle and the paths considered are portions of a ray inside a prism, the angles made with the third face by the ray before and after the double reflection are equal, and the incident and emergent angles in air are also equal. The essential condition is therefore that the large angle of the prism should be a right angle, the equality of the other angles being unimportant. The fact that in considering changes of direction only the projection in a principal plane need be considered enables a simple geometrical con-

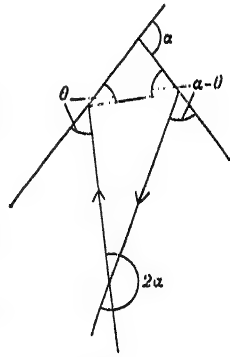


Fig. 4.

struction to be used, in which the path of a ray is represented by a straight line. Let ABC (*Fig.* 5) represent a principal section of a prism, and PQ a portion of the path of a ray within the prism which meets the sides in the order AC, AB, BC, AC. Let ABC′ be a triangle on AB, such that AB divides AC′BCA

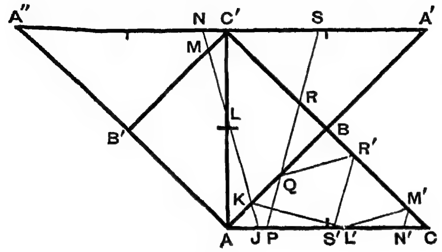


FIG. 5.

into two symmetrical triangles with AC′=AC, BC′=BC. Similarly on BC′ let A′BC′ be constructed symmetrically with ABC′. It is then easy to see that the actual ray path PQR′S′ corresponds to the extended straight line PQRS. The prism will produce no dispersion if C′A′ is parallel to AC, that is, if $2A+2C=\pi$, or if B is a right angle, and the prism is exactly equivalent to a plane parallel block of glass whose thickness is equal to the hypotenuse AC of the prism.

The prism is mounted in such a way that incident rays can only enter the face AC in the half nearer to A, and only rays which emerge in the half nearer C can proceed further through the instrument. Evidently, in addition to such a ray as PQRS, it is possible for a ray like JKLMN, which meets the sides in the order AC, AB, AC, BC, AC, to emerge, and the final direction may be parallel to that of R′S′, though the incident rays made angles of opposite signs with the normal to AC. Such rays evidently cause a laterally inverted image [1] to be superposed upon that previously considered, and steps must be taken to prevent this. If the angle made by the ray with the normal to AC is not small, the ray will meet AB and B′C′ at small inclinations, and the light which eventually reaches A″C′ will be too faint to be troublesome. The rays which make a small angle with the normal may be stopped by cutting a slit symmetrically across AC parallel to the line of intersection of AB and BC. From the images of this slit in the diagram it is evident that this simple device is adequate to prevent this "ghost" image, and that its only direct effect is to necessitate a somewhat larger prism to transmit the useful light than would otherwise be the case.

[1] It may be noted that the character of the image in this respect depends upon the number of reflections being even or odd.

The size of the prism is evidently important. The larger the prism the longer the ray path in the prism, so that variation in the size involves variation in the image position, and for a manufactured instrument the variation in size must be confined within narrow limits. The shaping of the base to fit in its seating is quite a distinct question. There will, in practice, be a limit to the permissible difference in the base angles, as a large inequality would involve a loss of light owing to the vertex of the prism, which is truncated, intruding into the region in which rays travel. These considerations, together with the fact that variations from a right angle in the vertical angle are only important on account of the prismatic effect of an error, lead to the conclusion that very great accuracy in the angles is not essential. Errors of the order of a minute of angle will be permissible: [2] this value may be contrasted with an accuracy of about a second necessary in a right angle when light may be incident first on either of the planes forming the angle.

If one of the prisms of a pair, situated in the way necessary to convert an inverting into an erecting telescope, is rotated so that its principal plane is brought into parallelism with that of the other prism, there is in one direction no inversion at all, and that in the other direction is a reinversion of the first inversion; in other words, no inversion has been effected. It may be inferred that the image rotates through twice the angle described by the prism. If, then, the principal planes of the two prisms are not at right angles to one another, the image of a vertical object will be inclined to the vertical. It is therefore necessary for the ridge of each prism to be very closely perpendicular to the longer axis of the base. The same principle shows that pyramidal error in one prism produces the same effect as an error in the right angle of the other prism. Advantage is sometimes taken of this effect to combine together a pair of prisms which tend to compensate one another's defects when these are too great to allow of the use of one of them with a normal prism.

It has already been seen that the prisms must be mounted in a way which prevents an alteration in the direction of their principal planes. When the prisms are placed between the objective and eyepiece it is equally important to prevent a translation of a prism in the direction of its length. The figure (*Fig.* 6) shows that the effect of such a movement is to displace the emergent rays through twice the distance through which the prism is moved. As a consequence of such a

[2] Tolerances of three minutes for the right angle, five minutes for the 45° angles, and three minutes for the pyramidal error have been proposed for service instruments. See *Report of the Committee on the Standardisation of Optical Instruments*, H.M.S.O.

displacement, rays from a point on the axis of the objective, which should emerge parallel to the axis, will form an image off the axis of the eyepiece, and the emergent rays will be

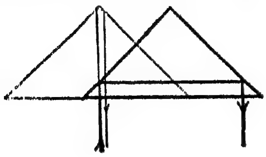


FIG. 6.

deviated. Thus rays through the two telescopes of a binocular, which should emerge in parallel directions, will be inclined to one another, and unless this inclination is small the instrument cannot be used, particularly for continuous observation, without causing discomfort to the observer. One of the most important details in a prismatic binocular is thus the means taken to ensure that the prisms are held immovable in their seatings. The method must not involve local strain in the prism, as this is liable to cause fracture, in addition to deterioration in the image. To secure correct adjustment within suitable limits the objective is usually mounted in a cell carried by two eccentric rings, which can be locked in any required azimuths. This device is illustrated in *Fig.* 7, where the small circle represents the area to any point of which the centre of the objective can be brought, as is desired.

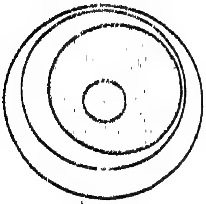


FIG. 7.

The error resulting from inaccurate adjustment of the instrument is of the same kind as would be produced by the use of a direct vision binocular in which the axes of the two telescopes were inclined to one another. It has thus been common to describe the error as "inclination of the axes," though this explanation is not strictly correct. The tolerable limits of error naturally refer to the inclinations of the two emergent beams, and are independent of the magnifying power. As would be expected, a much greater error can be allowed when the beams diverge on leaving the instrument, so that the eyes are required to converge, than when the beams converge, or have a relative inclination in the vertical plane. If the permissible errors of the emerging beam are translated into inclination of the axes, the amount of error allowable depends upon the power, but below a power of 8 it is customary to have fixed limits of error for the axes, as the attainment of the more stringent standard for low powers offers no difficulty. These axis errors may reach three minutes in the vertical plane and six minutes in the horizontal without causing discomfort to the observer.

Prismatic binoculars are generally mounted so that the two telescopes may rotate about an intermediate hinge, thus permitting the separation of the eyepieces to be varied from about 55 to 70 mm. If the optical axes of the telescopes are not parallel to the axis of the hinge, the inclination of the telescope axes will vary with the interocular distance. In testing binoculars it is accordingly customary to measure the errors of want of parallelism for three distinct interocular values.

A further adjustment of importance in any instrument in which the two eyes are used simultaneously is a means of focussing each telescope independently. The most satisfactory method is the direct one of mounting each eyepiece in a cell which may be moved in or out by a screw of very coarse pitch.

This very brief description of the principal adjustments which are necessary in a prismatic binocular suffices to indicate the complexity of the instrument in assembling its

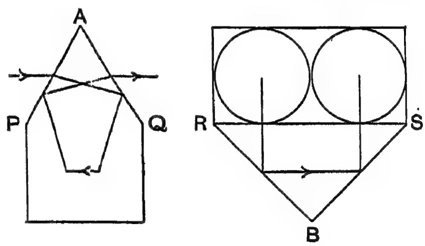


FIG. 8.

parts. In many designs attempts are made to avoid this by the use of external prisms. Other forms are illustrated in the article on German telescopes.[1] Another design is shown in the accompanying figure (*Fig.* 8), taken from a paper by H. Dennis Taylor.[2] The lower portion of the prism is of the "roof" type. The chief advantages of this prism are its compactness and the need for a single prism only. From the figures of its development (*Fig.* 9) it is readily seen that the roof angle B must be very closely a right angle, since parallel rays of the same beam must continue parallel whether their final positions are determined by developing the prism about RB and S′B or about SB and R′B. The development in the perpendicular plane shows that the bisector of the angle A must be perpendicular to the roof ridge.

[1] See "Telescopes of German Design."
[2] *Transactions of the Optical Society*, xxii. 63.

If, as in the illustration (*Fig.* 8), the prism is so enlarged to give increased stereoscopic power that the projections of the entrance and exit pupils do not overlap, the roof prism is used as an ordinary right-angled prism, and the accuracy necessary is not greater than in the

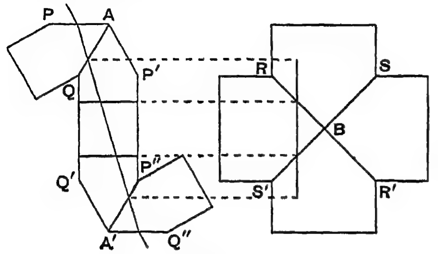


FIG. 9.

right angles of the prisms of the ordinary type of binocular.

§ (6) ANASTIGMATIC TELESCOPES.—In the paper just mentioned, Taylor describes a telescope in which the curvature and astigmatism of the image are corrected simultaneously. This implies the satisfaction of the Petzval condition, and therefore the presence of negative lenses in the instrument other than those used to remove chromatic aberration. The accompanying figure (*Fig.* 10) shows a section of the instrument taken from the paper. The objective is of approximately normal type. In front of the real image is placed a very powerful system of negative lenses N of fluor crown glass, achromatised with an extra dense flint. This lens is essential for the attainment of the desired correction, and forms an image of distant objects approximately in the plane of the field stop *dd*. The use of the negative lens in this position causes a large real image to be formed, corresponding to a focal length appreciably greater than the distance of the objective O from *dd*. The field lens F is necessarily a powerful lens to deviate to the eye lenses E_1 and E_2 the diverging beams which form the real image. F is made of dense barium crown glass to secure a high refractive index with a low dispersion. E_1 and E_2 are placed close together, separate lenses being necessary to attain the desired aperture ratio. The positive components are of light baryta flint glass and the negatives of extra dense flint. The corrections obtained resemble those of the most perfect photographic lenses, and the apparent field reaches 56 degrees. The telescope illustrated is of magnifying power 10, but systems of lower powers have been constructed. Two such telescopes may be mounted as a binocular by the addition of two of the prisms illustrated in *Fig.* 8, in front of the two objectives.

The chief difficulty in the construction of telescopes having their fields corrected in this way lies in the limited range of glasses, which necessitates the use of decidedly large curvatures in the component lenses, unless the number of lenses is much larger than is usual in normal designs. In theory there is no difficulty, as may be shown by considering a particular type of construction. As is well known, achromatic lenses can be made to yield images corrected for spherical aberration and coma for an object occupying a definite position. If in a train of lenses corrected in this manner the image formed by each lens occupies for the next succeeding glass the position of the object for which the latter is corrected, it is evident that the image produced by the entire instrument will be free from spherical aberration and coma. Under these conditions the curvature of the primary and secondary images produced by each individual lens will be less than that of the corresponding objects by the amounts $\kappa(3+\varpi)$ and $\kappa(1+\varpi)$ respectively, where κ is the paraxial power of the component and ϖ is its Petzval coefficient. It follows that the conditions for the simultaneous removal of curvature and astigmatism are

$$\Sigma\kappa=0 \text{ and } \Sigma\kappa\varpi=0,$$

the sums being extended to include every component lens. In practice these conditions

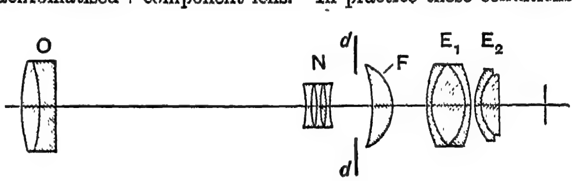


FIG. 10.

will need slight modification owing to the presence of higher order aberrations which must be partially compensated by those of lower order. The important feature of this particular solution is that the position in which the negative lenses occur in the system is immaterial, so it is at once evident that a solution is possible in which they are placed near the real image, the one position in which the divergence they tend to produce will not be an insuperable obstacle in the way of controlling the positions of the rays on emergence.

§ (7) NUMERICAL RESULTS.—The following table gives the leading dimensions and properties of some of the best-known patterns of telescopes and binoculars.

Type.	Magnifying Power.	Diameter of Objective in mm.	Diameter of Emergent Pencil in mm.	Angular Field Degrees.	Apparent Angular Field Degrees.	Length in cm.
Galilean binocular .	4	50	12	4	16	10
Prismatic binocular .	6	30	5	8	50	10
Prismatic binocular .	8	25	3	6	50	10
Terrestrial telescope .	45	75	1·7	0·8	35	110
Terrestrial telescope .	15	30	2	2·3	35	60
Gun-sighting telescope	8	50	6	5	40	60
Variable power gun-sighting telescope .	5 to 21	50	10 to 2·5	8 to 2	40	75

T. S.

TELESCOPES OF GERMAN DESIGN

THE general features of the optical systems employed in German telescopes are very similar to those which are to be found in British instruments.[1] The German designer, however, seems to make more use of specially designed prism systems to replace separated lenses for erecting the field of view. This characteristic, which leads to greater compactness, may be illustrated by a brief description of two German telescopic systems which the writer had occasion to dissect during the war.[2] The two instruments in question are naval gun-sighting telescopes, manufactured by Carl Zeiss, Jena.

§ (1) PERISCOPIC SIGHT.—A vertical section of the optical system is shown in *Fig.* 1. The special feature of the system is the prism D, which is also shown in perspective in *Fig.* 1A. It is of the "roof" prism type, and is designed to displace the optical axis of the instrument parallel to itself through a distance of 5·5 cm. and to reverse the field. The light enters at the face *a* (*Fig.* 1A), and is totally reflected internally at *b* and then at *c*. The light from one half of the field is next reflected from the face *d* to the face *e*, that from the other half of the field being reflected from *e* to *d*. The light then passes out of the prism at the face *c*. The path of the central ray is indicated in *Fig.* 1A by means of arrow-heads. In order to prevent overlapping or separation of the two halves of the field the angle between the faces *d* and *e*, which form the "roof" portion of the prism, must not differ from 90° by more than a few seconds.

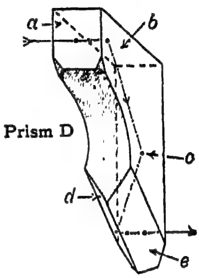


FIG. 1A.

The angles between the other faces of the prism are approximately as follow:

Angle between *a* and *b*, 60°
" " *b* and *c*, 120°
" " *c* and *d*, 52° 45′
" " *d* and *e*, 52° 45′

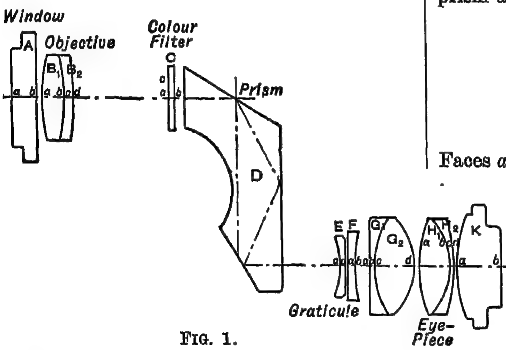


FIG. 1.

Faces *a* and *c* are parallel. This type of prism is also employed in many British instruments, particularly in dial sights, but the Germans seem to use it much more frequently.

§ (2) RIGHT ANGLE SIGHT.—A vertical section of the optical system is given in *Fig.* 2. The instrument is designed to deflect the light through a right angle, and is evidently intended for use in a position where the optical axis of the eyepiece makes an angle of 45° with the vertical, the eyepiece being directed upwards. The prism D′ in this instrument is of rather an unusual type. It is shown in perspective in *Fig.* 2A. The light enters at the face *a*′, is reflected internally at *b*′, and then at *c*′, and passes out of the prism at the face *d*′. The path of the central ray is indicated by means of arrowheads.

[1] See article on "Telescopes."
[2] For a complete description of the instruments see J. S. Anderson and A. B. Dale, *Opt. Soc. Trans.*, 1919, xx. 315.

The angles between the faces of the prism are approximately as follow :

Angle between a' and b', 45°
" " b' and c', 75°
" " a' and c', 90°
" " c' and d', 67° 30′
" " a' and d', 90°

These instruments also serve to illustrate a rather novel method of illuminating the cross-lines for night work. The cross-lines are engraved on the flat surface a of the graticule lens (F in *Fig.* 1 and F′ in *Fig.* 2), the edge of which is polished. On the outside of this lens there is mounted a cell, in a groove of which are two small glass tubes of circular cross-section and bent round in the form of rings. These tubes are filled with a fine powder which consists of a luminous compound having approximately the same luminosity as that given by the British ·2 mg. radium bromide per gr. compound. The light from this compound serves to illuminate the cross-lines when the instruments are being used in the dark. If the light should prove to have a disturbing effect in finding an object at night the illumination can be screened off by moving the cell, which contains the tubes, a short distance along the axis. This is accomplished by means of a worm-gear arrangement.

For more detailed descriptions of German telescopic systems reference may be made to *The Theory of Modern Optical Instruments*, by Dr. Alexander Gleichen, translated by H. H. Emsley, B.Sc., and W. Swaine, B.Sc.

J. S. A.

Temperament : a term used in music to denote a system of musical notes in which certain intervals are purposely modified in order to meet the requirements of practical convenience and allow of free modulation with fewer notes. See "Sound," §§ (4), (5).

Temperament, Equal : the musical temperament most commonly used, in which the semitones, whether diatonic or chromatic, are exactly half the tones. See "Sound," § (6) (iv.).

Attained on wind-instruments with special valves. See *ibid.* § (40).

Temperaments, Chief Musical, and Just Intonation, tabulated. See "Sound," § (6) (v.), Table I.

Test of Mirror of Reflecting Telescope. See "Telescope," § (11).

Test-case of Trial Lenses : a series of the lenses forming the various possible combinations of spectacle lenses, used in testing a patient's vision. See "Ophthalmic Optical Apparatus," § (9).

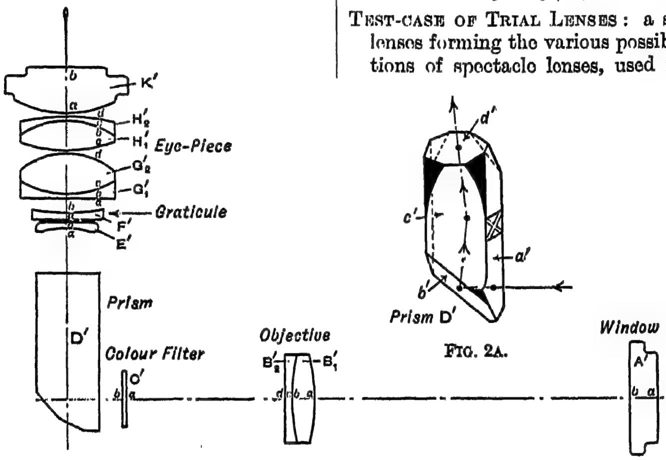


Fig. 2a.

Fig. 2.

Test Plate : an appliance of the practical optician having a very perfect optical surface, either flat or curved. See "Optical Parts, The Working of," § (4); also "Interferometers, Technical Applications of."

Tetra-iodo Fluorescein : a reagent for testing optical glass. See "Glass, Chemical Decomposition of," § (3) (i.).

THEODOLITE, THE

§ (1) The theodolite is not only the most important of surveying instruments, but one of the oldest. Leonard Digges in 1571 [1] describes a "Theodelitus" consisting of a graduated horizontal circle, with an "Alidade" or sight pivoted on a vertical axis through the

[1] Leonard Digges, *A Geometrical Practical Treatise, etc.*, 1571.

centre of the circle. The theodolite was therefore primarily an instrument for measuring horizontal angles, *i.e.* the angles between planes containing the vertical axis of the instrument and the various objects observed

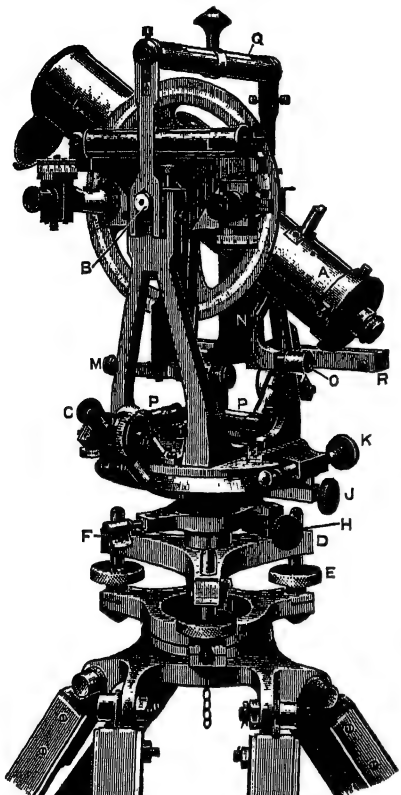


FIG. 1.

to. Digges also shows how a "quadrant" can be added so as to observe angles of elevation or depression at the same time as the horizontal angles. The theodolite as now made is generally provided with a vertical arc, and most modern instruments are made so that the sighting telescope will "transit" on its horizontal axis, *i.e.* so that it can be reversed end for end without removing it from its bearings; in this case the vertical arc (if provided) is generally a complete circle. Such instruments are known as "transit theodolites," or shortly (especially in the U.S.A.) as "transits." The instrument was first developed as an accurate portable instrument by English makers in the latter part of the eighteenth century. The 3 ft. theodolite by Ramsden made in 1787 for the Royal Society was the first which was sufficiently accurate to detect the spherical excess of terrestrial triangles. For the history of the theodolite, see Laussedat.[1]

Instruments are now made with circles ranging from 3 in. in diameter (for explorers) to 12 in. (for geodetic work).

§ (2) DESCRIPTION. — *Fig.* 1 illustrates a 6-in. micrometer transit theodolite by Troughton & Simms, and *Fig.* 2 a 5-in. example by E. R. Watts & Sons. In both cases the circles are read by micrometer microscopes to 10 seconds, and to single seconds by estimation.

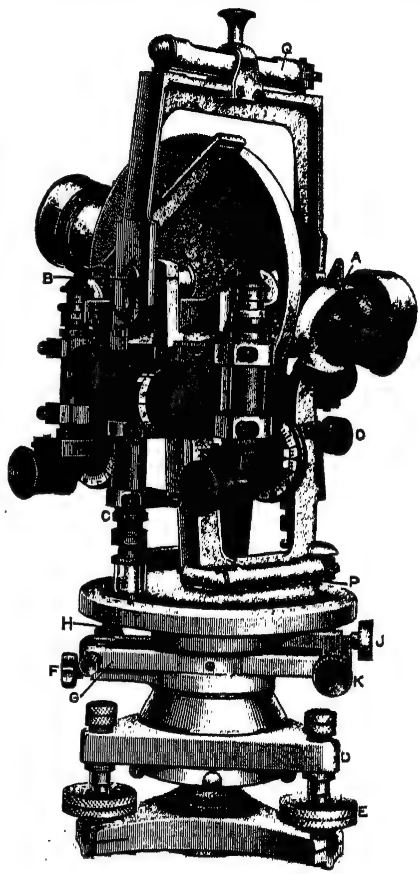


FIG. 2.

The telescope A is an ordinary astronomical telescope provided with cross hairs, and mounted on a horizontal axis B. When the instrument is in perfect adjustment, the collimation line of the telescope intersects the horizontal axis at right angles. The diaphragm can be adjusted by the screws

[1] A. Laussedat, *Recherches sur les instruments, les méthodes, et le dessin topographiques*, 1898.

shown in *Fig.* 1 (in *Fig.* 2 they are concealed by a screw-on cap). The axis B is carried by arms attached to the upper plate of the instrument. The upper plate is shaped to protect the divided circle, and carries the microscopes for reading the circle. The whole instrument is supported on a trybrach D, with three levelling screws E. The upper plate and the lower plate (carrying the divided circle) are capable of rotation independently on concentric vertical axes. A clamping screw F clamps the lower plate to the trybrach in such a manner that a small relative motion can be given by the slow motion tangent screw H. Similarly clamping screw J and tangent screw K connect the upper and lower plates. The vertical axes should intersect the horizontal axis at right angles, and at the same point in which the collimation line intersects it. An adjustment is provided for raising or lowering one end of axis B so as to make it perpendicular to the vertical axis. The vertical circle is rigidly attached to the telescope. In *Fig.* 1 the microscopes for the vertical circle are attached to an arm which also carries the sensitive spirit level L, and which is mounted rotatably on the horizontal axis, and prevented from turning with the telescope by the clip screws M. In this instrument the vertical circle with its microscopes and level is lifted out of its bearings for packing in its transport case; the vertical circle is unprotected. In the instrument shown in *Fig.* 2, the axis B revolves in segmental bearings, and is not intended to be removed from them, the microscopes and bubble L are fixed to the arms supporting the axis B, the bubble is viewed by means of a mirror, and the vertical circle is completely enclosed. When the vertical circle is read by verniers, these are always carried on an arm, as are the microscopes in *Fig.* 1. A clamping screw (not visible in either photograph, but visible in *Fig.* 3) clamps the telescope axis to the arm N, while a tangent screw O provides a slow motion. When in proper adjustment, the microscopes should read zero when the bubble L is levelled and the telescope is horizontal. One or two levels P on the upper plate provide for levelling the vertical axis. A striding level Q rests on the ends of axis B to ensure its being horizontal; this lifts off for packing, or when not required.

A diagonal eyepiece is provided for taking sights at high angles of elevation. Some instruments have a circular compass on the upper plate, while others have a trough compass that can be attached to one of the arms (R, *Fig.* 1). In some cases the main spirit level is attached to the telescope instead of to the microscope arm; this is more convenient if the instrument is to be used as a level, but less so for reading accurate vertical angles.

§ (3) Details of Construction.—These vary considerably in different instruments. *Fig.* 3 shows details of an instrument similar to *Fig.* 1. The upper plate is fixed to a coned spindle S, which revolves in the hollow spindle T carrying the circle, which again revolves in a bearing fixed in D. A nut fixed to the bottom of S prevents the withdrawal of the spindles. In small instruments the weight is sometimes taken by the coned bearings, but in larger instruments only sufficient weight should be so taken to prevent shake. *Fig.* 4 shows one method of arranging this. In order to prevent shake and take up wear in the levelling screws E, the ends of the trybrach arms are often split vertically and the sides pulled together against the screw E by the screw U (*Fig.* 3). In older instruments the thread of screws E was generally left exposed, and the wear due to dust was often considerable; in *Figs.* 1 and 3 they are shown covered in. *Fig.* 5 shows another method of arranging this; a nut V is screwed and fixed into D, and the footscrew works inside V. The upper portion of V is split, and its outer surface is coned; a dust cap W is coned inside, so that when screwed down it grips the top of V and presses it on to E.

Fig. 6 shows the construction of the lower part of *Fig.* 2. In this case, S is a hollow cylinder of hardened steel, which revolves in a fixed bearing Y, while the circle is carried by a hollow bearing T, which revolves round S. The whole weight is taken by two ball-bearings. This makes a very easy movement, and wear of the main axes is practically eliminated; a further advantage of this construction is that there is no difficulty in making the two vertical axes coincident. The hollow spindle enables the telescope to be used for centring the instrument over the station mark. A nut X prevents the spindle from being withdrawn. The whole support Y of the upper portion of the instrument has a small horizontal movement with respect to D for fine adjustment of the centring, being clamped by the screw Z; *Fig.* 3 shows a similar movement below the levelling screws.

In large instruments the double-coned bearing is not used; the lower plate is fixed direct to the base of the instrument, and, though capable of rotation, it is not provided with any slow motion.

The construction is simplified and cheapened if verniers are substituted for microscopes, but at the expense of the accuracy of reading the angles, and verniers may often be used with advantage on the vertical circles. Great accuracy of reading vertical angles is seldom required, except for certain astronomical observations, and if the theodolite is fitted with arrangements for using Talcott's method, all necessary astronomical observations for

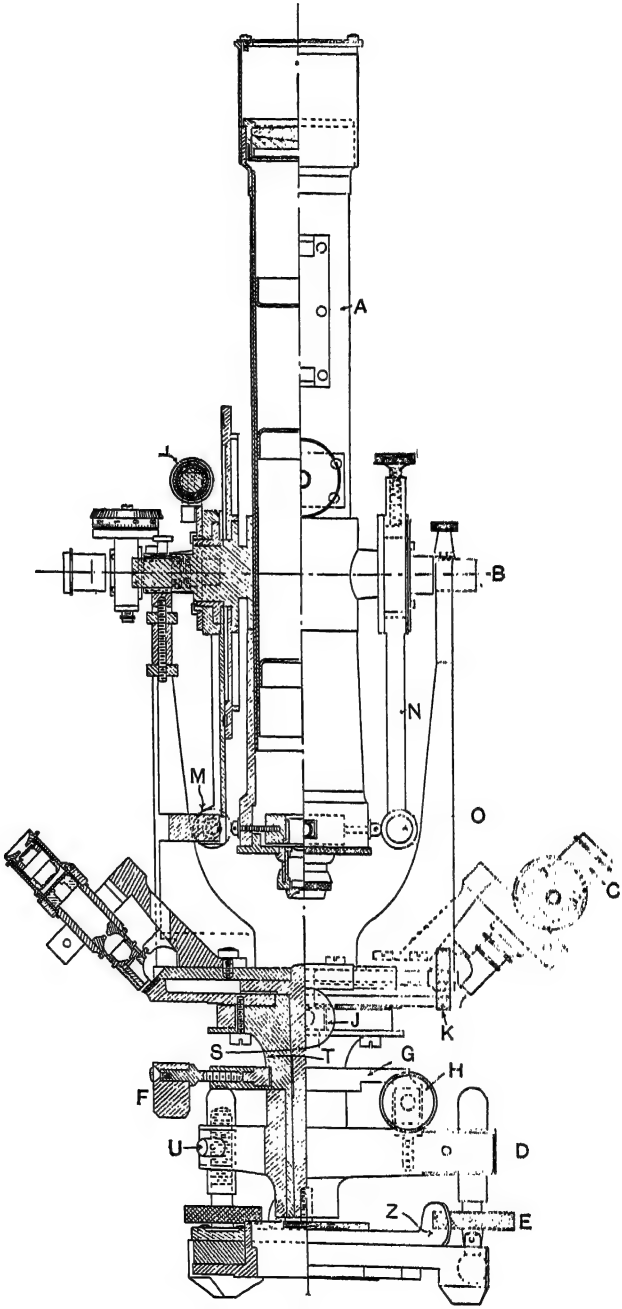


Fig. 3.

survey work can be carried out without any accurate readings on the vertical circle. See "Surveying," § (27).

§ (4) METHOD OF TAKING OBSERVATIONS.—The instrument must be set up and carefully centred over the station mark by means of a

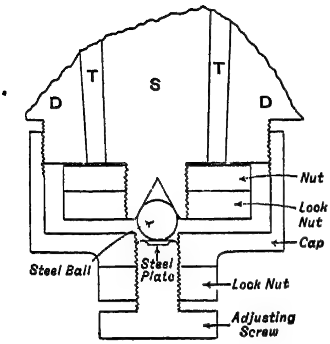


FIG. 4.

plumb-bob, or by a nadiral telescope, and the vertical axis brought truly vertical; for accurate work the most sensitive bubble available should be used for the final levelling. The eyepiece is focussed on the cross hairs and then the telescope focussed on the distant object. It should be noted that the eyepiece focussing depends on the sight of the observer and the telescope focussing on the distance of the object; the image formed by the object glass must be in the plane of the cross hairs, or the cross hairs will appear to move relative to the distant object if the observer's eye moves.

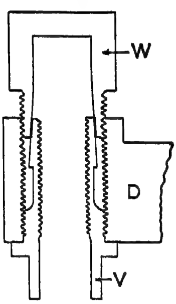


FIG. 5.

(i.) *Observing Horizontal Angles.*—The usual method is known as the direction method. The lower plate is clamped, and the telescope is turned on the first object, the object being intersected by the vertical cross hair by means of the slow motion screw. The horizontal circle is then read. The upper plate is then unclamped and the telescope swung to the right till the next object appears in the field of view, the plate is clamped and the final intersection carried out by the slow motion screw, and the circle again read. All stations are intersected in the same manner, swinging to the right, and closing on the original mark as a check. The telescope is then transited and a similar round taken swinging left. Observations taken with the vertical circle on the left and right of the telescope are known as "face left" and "face right" respectively. In all accurate work an equal number of rounds should be taken "swing right" and "swing left" and an equal number "face right" and "face left." By taking sets of rounds on different zeros, *i.e.* with different readings on the circle for the pointing on the first mark, the errors of graduation are averaged. In geodetic work at least 6 zeros are used, with two "faces" on each zero and two "swings" on each face. In tertiary work 2 zeros, 2 faces, and 1 swing on each face may suffice. The number of rounds taken depends on the accuracy required, and the accuracy attained should be judged by the closing errors of the

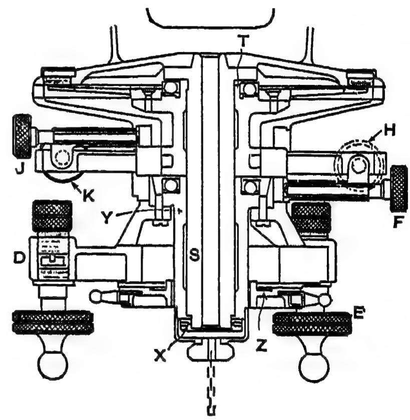


FIG. 6.

triangles rather than by the agreement of the individual measurements of the angles.

A method sometimes used is the "repetition" method, when each angle is measured separately. The one object is intersected as before and the circle read, the upper plate is then unclamped, the telescope turned on the second object, which is then intersected; leaving the upper plate clamped the lower is unclamped and the telescope turned on the first object, which is then intersected with the lower tangent screw; the whole process is repeated several times; the final angle read, divided by the number of repetitions, gives the actual angle. A similar set should then be taken on the other face. This method reduces the error due to reading the circle, and is sometimes of use, but with modern instruments the errors due to the circle readings are likely to be less than the errors due to movement of the clamps, and the method is seldom used. It is never used for geodetic work, and large theodolites are seldom provided with the necessary slow motion on the lower plate.

Some large instruments are provided with a "watch" telescope attached to the lower plate. A reference object is intersected by this and observed at intervals, with the result that any movement of the lower plate, or twist of the stand, can be observed and allowed for.

Horizontal angles are best observed with the horizontal axis unclamped.

(ii.) *Observing Vertical Angles.*—Vertical angles are best observed with the upper plate unclamped. The instrument must be very carefully levelled. The position of the bubble on the microscope arm should be noted at the time of observation, and a correction made if it is not in the centre of its run. A second reading should be made on the opposite face; this is even more important in the case of vertical angles than in the case of horizontal angles, and a reading on a single face is of little value owing to collimation errors.

§ (5) Errors and Adjustments. (i.) *Eccentricity of Circle.*—It is easily shown that if there are two or more verniers (or microscopes) evenly spaced round the circle the error due to any eccentricity of the circle with respect to its axis cancels out if the mean reading of the verniers is taken, and that the same applies for all practical purposes if the verniers are not quite exactly spaced. A small adjustment can be made if necessary in the spacing of microscopes by moving the drum and the comb, but the spacing of the verniers and the eccentricity of the circle are matters for the makers.

(ii.) *Horizontal Circle not in Plane Perpendicular to the Axis.*—The maximum error introduced is about $\frac{1}{4}a^2$, where a is the (small) angle of tilt; this error is negligible, as if a were large enough to produce any appreciable error the verniers or microscopes would be noticeably untrue on the circle.

(iii.) *Vertical Axis Vertical, but Horizontal Axis not Horizontal.*—If the horizontal axis make an angle ν with the horizontal the error introduced is $\nu \tan h$, where h is the angle of elevation of the angle observed. This error is especially important when horizontal angles are measured to points at a considerable elevation, as in astronomical observations for azimuth. The error will cancel out in the mean if observations are made on both faces, or the striding level can be used and a correction made for any dislevelment found at the moment of observation.

(iv.) *Vertical Axis not Vertical, but Horizontal Axis Perpendicular to Vertical Axis.*—The error depends on the angle a between the direction observed and the plane in which the axis is tilted, the angle of tilt β, and the elevation of the object observed h, and amounts to $\beta \sin a \tan h$. As $\beta \sin a$ will represent the actual tilt of the horizontal axis its amount can be ascertained by the striding level and a correction made if necessary.

It will be noticed that this error due to dislevelment cannot be eliminated by any method of observation, but that they only become of importance when points are being observed to at a considerable angle of elevation.

(v.) *Collimation Line not Perpendicular to the Horizontal Axis.*—In this case the line of sight traces out a small circle parallel to the great circle through the zenith. If u is the angle the collimation line makes with the perpendicular to the horizontal axis, and h the angle of elevation, then the error is $u \sec h$. If a second observation be taken after changing face the error will remain the same in magnitude but will be of opposite sign, hence the mean of the two observations will be free from this error. The collimation error can be ascertained in several ways; the two commonest are: (*a*) Intersect a distant object, nearly at the same level as the instrument, leave the horizontal circle clamped, lift the telescope out of its pivots and replace after turning through 180° about its length; if the object be still intersected the instrument is correctly collimated in azimuth, but if not the actual error is half the apparent error, and can be corrected by moving the cross hairs. This method is only possible when the telescope can be lifted out and reversed. (*b*) Intersect an object as before, transit the telescope, and revolve the upper plate till the object is again intersected. If the difference of the two readings on the horizontal circle is exactly 180° the instrument is in collimation, if not half the difference is the collimation error.

(vi.) *Collimation Line Perpendicular to the Horizontal Axis, but Distant d from the Vertical Axis.*—In this case the error is ϵ, where $\sin \epsilon = d/l$, l being the distance of the object from the instrument. This error also cancels out in the mean of two observations on different faces, and in any case can only be appreciable on short sights.

(vii.) *Vertical Cross Hair not Vertical.*—This is tested by intersecting an object, and moving the telescope in altitude, when the object should remain intersected over the whole length of the hair. If it does not, the diaphragm must be rotated till it does. Any residual error can be avoided by always intersecting on the same portion of the cross hair.

The above refer to horizontal angles, but errors (i.), (ii.), (vi.), and (vii.) apply *mutatis mutandis* to vertical angles, and are eliminated in the same way.

(viii.) *Vertical Collimation.*—All vertical angles are read with reference to a spirit-level, and when the instrument is in adjustment

and with the vernier (or micrometer) arm bubble in the centre of its run, and the verniers reading zero, the collimation line should be horizontal. This is tested by reading the elevation of any distant object on both faces, making any necessary correction for any movement of the bubble between the two observations. The elevations read should be the same; if not, the true elevation is the mean of the two readings, and the error is half the difference. The error can be removed by adjusting the bubble, by adjusting the verniers (or microscopes), or by altering the cross hairs. The last method is not good, as the cross hairs should be on the axis of the object glass; the choice between the other two methods depends on the construction of the instrument. In instruments such as those illustrated in *Figs.* 1 and 3, where the bubble is attached to the arm carrying the verniers, the true elevation of a point is found by readings on two faces. The verniers are then set to the correct reading by means of the clip screws; and, lastly, the bubble is brought back to the centre of its run by means of its adjusting screws. In instruments such as that illustrated in *Fig.* 2, where the bubble is attached to the upper plate, it must be set with reference to the vertical axis, and in this case the micrometers must be made to read correct by altering the drum, and (if necessary) the comb, or even by moving the whole microscope.

All the errors considered above, except the levelling errors, can be eliminated by suitable means of observation, but it is still desirable that the errors themselves should be small. It is sometimes only possible to observe on one face, and it is sometimes desired to pick up an indistinct object, or faint star, by setting the known angles on the circles. There remain four important errors:

(ix.) *Errors of Intersection.*—These depend on the skill of the observer, the power and optical qualities of the telescope, the character and illumination of the object intersected, the suitability of the graticule, and the steadiness and clearness of the atmosphere. In any given circumstances the error can only be reduced by taking the mean of a number of observations. It is sometimes advantageous to use a micrometer eyepiece, and obtain extra intersections by its means.

(x.) *Errors of Graduation.*—These depend on the maker of the instrument, and the best modern instruments should have no error greater than a few seconds, and in the larger instruments the maximum error should not exceed two or three seconds. The resulting errors can be reduced by taking a number of readings on different portions of the circle, *i.e.* by taking rounds of angles on different zeros.

(xi.) *Errors of Reading the Circles.*—This depends on the method of reading adopted (see article "Divided Circles"), the micrometer microscope being the most accurate. In the case of micrometer microscopes the error depends very largely on the illumination of the circle. The error can be reduced by taking a number of readings.

(xii.) *Atmospheric Refraction.*—Horizontal refraction [1] is seldom appreciable, unless the conditions are obviously unsuited for observations, or unless the line of sight grazes a building or the side of a hill; a rather noticeable case is given in the *Professional Papers of the Ordnance Survey*, No. 2, p. 11. Vertical refraction is always important.

E. O. H.

THERMAL ENDURANCE OF GLASS. See "Glass," § (3').

THERMOELECTRIC METHOD OF SPECTROPHOTOMETRY: the method depending on the use of a thermopile and galvanometer. See "Spectrophotometry," § (16).

THERMOPILE AS A PHYSICAL PHOTOMETER. See "Photometry and Illumination," § (35).

THOMPSON - STARLING PHOTOMETER. See "Photometry and Illumination," § (27).

THORIUM ACTIVE DEPOSIT, DECAY OF. See "Radioactivity," § (20) (i.).

THORIUM X: the first product of the radioactive disintegration of thorium. See "Radioactivity," § (3).

THRESHOLD: door, entrance, the place or point of entering; used in connection with the senses to indicate phenomena associated with very feeble stimuli just sufficient to produce a sensory response. See "Eye," §§ (4) and (14).

TIME, OBSERVATION FOR, for survey purposes. See "Surveying and Surveying Instruments," § (25).

TIME EFFECTS, VISUAL. See "Eye," § (17).

TRANSFORMATIONS OF RADIOACTIVE SUBSTANCES, THEORY OF SUCCESSIVE. See "Radioactivity," § (20).

TRANSFORMER: an appliance for converting alternating current at low voltage to current at high voltage, or *vice versa*. See "Radiology," § (10).

TRANSIT MOUNTING FOR TELESCOPES. See "Telescope," § (14).

TRANSMISSION, SPECTRAL. See "Spectro photometry," § (14).

[1] See article "Trigonometrical Heights and Terrestrial Atmospheric Refraction," Vol. III.

TRANSMISSION LOSSES with carbon arc as source cannot be less than 40 per cent. See "Projection Apparatus," § (8).

TRANSPARENT SUBSTANCES, TESTS FOR HOMOGENEITY OF

IN making use of transparent substances, such as glass, quartz, fluorite, etc., for optical purposes it is essential to know the degree of homogeneity of the pieces of material employed. The main defects met with take the form of bubbles and striae or veins due to the presence of layers having different refractive indices. Bubbles can usually be detected by the naked eye or with the help of a microscope. Want of homogeneity may be revealed by using either of the two following methods.

§ (1) SHADOW METHOD.—This can conveniently be carried out on an ordinary spectroscope. The telescope is moved into the position in which its axis is collinear with that of the collimator. The eyepiece is removed and a fine wire is mounted so as to coincide exactly with the image of the slit. If the slit is illuminated and one's eye is placed behind the wire, the field of view will appear dark. If a parallel-sided block of glass is inserted between the collimator and telescope objectives the presence of striae will be revealed by light streaks. If the specimen is in the form of a prism the telescope must be moved round until the refracted image of the slit coincides with the wire; in this case monochromatic light should be used. If the specimen has an irregular shape or has no polished surfaces it may be immersed in a liquid of the same refractive index (for a given wave-length) and examined as in the case of a parallel-sided block. A lens may be tested without using a spectroscope: an illuminated pin-hole is set up at some distance from the lens, and if one's eye is placed at the conjugate point, any striae that may be present will appear as more or less dark streaks on a bright background. The visibility of the striae is usually improved by obscuring the image of the pin-hole either partially or almost entirely with a diaphragm; in the latter case the striae will appear bright on a dark background.

§ (2) INTERFEROMETER METHOD.—A complete discussion of this method is given in the article on "Interferometers: Technical Applications," § (5).

§ (3) TEST FOR STRAIN IN OPTICAL GLASS.—The presence of strain in optical glass is usually due to imperfect annealing, and it may be detected by means of the double refraction which is caused by the strained condition. If a specimen is placed between crossed Nicol prisms the presence of internal strains will be revealed by the appearance of more or less bright patches in the field. A convenient arrangement for obtaining a large field of view is to allow a beam of light, diverging from an illuminated pin-hole, to pass through the polarising Nicol and fall on a piece of ground glass. The specimen is then examined against the ground glass background [1] by means of the analysing Nicol. In order to increase the sensitivity of the test a wave-plate may be inserted in front of the analyser and orientated so as to give the sensitive first order violet colour between crossed Nicols, when no specimen is present. The presence of double refraction in a specimen will then be revealed by a change of the violet colour to a tint of a lower or higher order.[2]

J. S. A.

TRAVERSES: Computing and Plotting of. See "Surveying and Surveying Instruments," § (14).
General. See *ibid.* § (11).
Use of tapes in. See *ibid.* § (12).

TRAVERSING. See "Surveying and Surveying Instruments," § (9) (ii.).

TRIAL-FRAME: a frame used to mount lenses when testing a patient's vision. See "Ophthalmic Optical Apparatus," § (8) (iv.).

TRIANGULATION, METHOD OF. See "Surveying and Surveying Instruments," §§ (9) (i.), (10).

TRICHROMATIC THEORY: a theory of colour vision. See "Eye," § (9).

TRIGONOMETRICAL METHODS applied to the tracing of rays through a series of coaxial spherical refracting surfaces. See "Optical Calculations," § (1).

TROMBONE: a brass wind-instrument, in which no restriction is placed on the intonation possible, the mechanism for the scale consisting of a U-slide, which may be drawn out so as to flatten the pitch continuously by any desired amount to six semitones. See "Sound," § (43).

TROTTER ILLUMINATION PHOTOMETER. See "Photometry and Illumination," § (56).

TROTTER PHOTOMETER. See "Photometry and Illumination," § (28).

TRUMPET, BACH: a brass wind-instrument with a tone of great nobility and brilliance. See "Sound," § (41).

TRUMPETS IN B♭, F, ETC. See "Sound," § (42).

TUBING: manufacture of Glass Tubing. See "Glass," § (18) (ii.).

[1] A certain amount of depolarisation takes place at the ground surface, but for qualitative examination this is not appreciable.

[2] Cf. *Opt. Soc. Trans.*, 1917, xviii. 88.

TUNGSTEN ARC LAMP

THIS lamp, which was developed by Gimingham and Mullard,[1] forms a very convenient source of radiation for many purposes. It consists essentially of an arc between tungsten electrodes, enclosed in a neutral atmosphere of nitrogen or argon, together with an auxiliary gear for striking the arc. Its simplest form is illustrated in *Fig.* 1. The ends of a tungsten filament BB

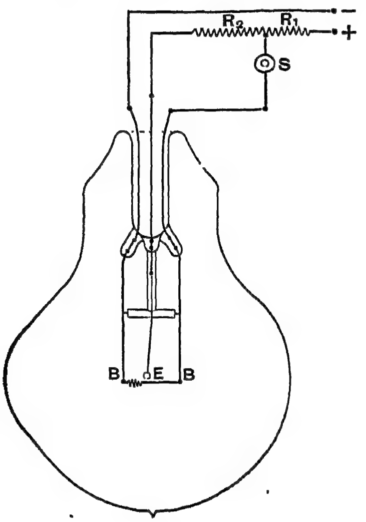


FIG. 1.

are connected through the push switch S and the resistance R_1 to a voltage supply. A bead of tungsten E, supported on a tungsten stalk, is also connected through the additional resistance R_2 to the positive pole of the supply. On pressing the switch S the filament BB is raised to incandescence and ionises the gas. When the switch is released, this enables the arc to strike between E and B. The arc is usually run with E at a temperature of the order of 2600° to 3000° abs., but it may be run up to the melting point of tungsten or, by reducing the pressure of the surrounding gas and increasing the size of the anode, at a dull red heat. In the latest form of lamp, the ioniser BB consists of tungsten wire partly straight and partly wound into a short spiral. On the straight portion is threaded a tube composed of a mixture of tungsten and certain refractory oxides, such as zirconia, yttria, or thoria. The arc first strikes on the uncovered spiral of tungsten wire and, when the electrodes warm up, passes on to the tubular portion which presents a shorter path (see *Fig.* 1). As the tube of refractory material is less liable to disintegration under the action of the arc than the tungsten wire, this device increases considerably the life of the lamp. Lamps of this type are made for candle-powers from 30 to 500 and for voltages of 100 and upwards.

For higher candle-powers a modified form of lamp has been evolved with an additional electrode of tungsten which is in the form of a square plate. The lamp has an ioniser and a bead of tungsten which are arranged exactly as in BB and E of *Fig.* 1. The additional plate of tungsten is fixed close to the bead and is connected to a double change-over switch, one position of which connects the volt supply to the ioniser and bead and the other position to the bead and plate. In the first position the polarity of the bead is positive and in the second negative. The arc is first struck between the ioniser and the bead by means of a press switch as described above, the change-over switch being in the first position. When the bead has become incandescent, the switch is thrown to the second position which enables the arc to strike between the bead and the plate. By the use of substantial electrodes, candle-powers up to 10,000 have been obtained, the size of anode for this value being 1 inch square.

The characteristics of the tungsten arc lamp are illustrated in *Fig.* 2. The normal working efficiency is taken as 0·5 watt per candle-power, and the percentage of normal voltage and candle-power as well as the watts per candle-power are shown for the lamp when run at efficiencies higher and lower than the normal. Curve A is similar to that of

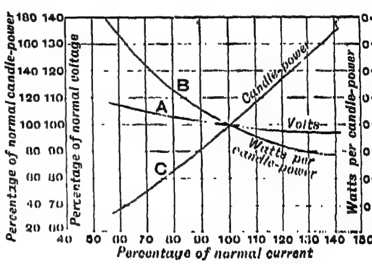


FIG. 2.

the ordinary carbon arc, but exhibits less instability. The volt drop across the arc steadily decreases with increase of current until the sputtering point is reached, when the pressure suddenly falls. The efficiency steadily rises with increasing current, reaching 0·3 watt per candle-power at the sputtering point. It should be mentioned that the figures for efficiency refer only to the energy dissipated in the lamp itself and take no account of that dissipated in the necessary

[1] *Journ. Inst. Elect. Eng.*, 1915, liv. 15.

ballast resistance. The intrinsic brilliancy at the normal working current is stated to be generally about 12,000 candle-power per square inch.

The distribution of the energy radiated from the tungsten arc in the visible spectrum does not differ appreciably from that of the tungsten filament lamp at the same temperature.[1]

The tungsten arc finds a number of useful applications where a source of light is required of small size and high intrinsic brilliancy. F. H. S.

[1] See "Incandescence Lamps," § (8), Vol. II.

— U —

Ulbricht Globe: a form of integrating photometer. See "Photometry and Illumination," § (47).

ULTRAMICROSCOPE, THE, AND ITS APPLICATIONS

§ (1) The Ultramicroscope.—The term "ultramicroscope" is apt to be rather misleading, as it may convey the idea that it is the name of an instrument with which one is able to see greater detail in an object than with an ordinary microscope. This, however, is not the case; the term should really be looked upon as applying to an instrument which reveals the presence of very minute particles of matter. The limit of resolution of a microscope—that is, the smallest distance between two points which can just be seen to be separated—has been theoretically deduced by Helmholtz, Abbe, and others,[1] who have shown that it depends on diffraction phenomena. The value of this smallest distance ϵ is given by

$$\epsilon=\frac{\lambda}{2a},$$

where λ is the wave-length of the light employed and a is the numerical aperture of the microscope. Now there are practical limits to the value of the numerical aperture, the maximum value obtainable with present-day objectives being about 1·53; it is not likely that this will be increased appreciably in the future. The other factor which determines the limit of resolution is the wave-length of the light used. By employing ultra-violet radiations one can reduce the value of ϵ; this has been done with success in processes of microphotography. One can see, however, that there are practical limits to the minuteness of detail that can be revealed by the microscope. On the other hand, there is theoretically no limit to the smallness of objects the presence of which can be revealed. If a sufficiently high intensity of illumination could be obtained, it should be possible to see a molecule, or rather to see the light scattered by a molecule, although it is quite impossible to differentiate molecules unless their separation is at least as great as the limiting value of ϵ. Faraday[2] was the first to look for an effect exerted on light by very fine metal particles suspended in liquid or solid media. Shortly afterwards Tyndall[3] showed how such particles could be seen by causing a beam of light to pass through the solution, the particles scattering the light in all directions. No advance was made in the systematic investigation of very minute particles until Zsigmondy turned his attention to the problem at the beginning of the present century. Together with Siedentopf, he improved the methods of observing such particles; they invented what they termed an "ultramicroscope," with which it is possible to render individual particles visible.[4] All that Faraday had been able to do was to reveal the presence of groups of particles. The important subject of colloidal chemistry, which deals with the study of colloidal particles—that is, very minute material particles in liquid or solid solution—has been developed largely as the result of Zsigmondy's pioneer work.

Since the invention of the ultramicroscope by Zsigmondy and Siedentopf a number of other forms of the instrument have been devised. The underlying principle of all types of ultramicroscope is the utilisation of dark-ground illumination. This method of illumination has long been employed in ordinary microscopy in connection with the study of such objects as diatoms. It consists essentially in illuminating an object in such a way that only the light which is scattered or diffused by the object enters the microscope, no direct light from the beam being allowed to enter the observer's eye. The different types of dark-ground illumination which are used in ultramicroscopy may be classed as follows:

(i.) *Orthogonal Illumination.*—This is the system utilised in the Zsigmondy and Sieden-

[1] Cf. E. Abbe, *Roy. Mic. Soc. Journ.*, 1881, p. 388; Lord Rayleigh, *Phil. Mag.*, 1896 (5), xlii. 167; *Roy. Mic. Soc. Journ.*, 1903, p. 447.

[2] M. Faraday, *Roy. Inst. Proc.*, 1854–58, ii. 310, 444; *Phil. Mag.*, 1857 (4), xiv. 401, 512; *Roy. Soc. Phil. Trans.*, 1857, p. 145.

[3] J. Tyndall, *Phil. Mag.*, 1869 (4), xxxvii. 384.

[4] R. Zsigmondy, *Zur Erkenntnis der Kolloide*, Jena, 1905; *Colloids and the Ultramicroscope*, trans. by J. Alexander, New York, 1909; H. Siedentopf, *Phys. Zeits.*, 1905, vi. 855; 1907, viii. 85; *Zeits. f. wiss. Mikroskopie*, 1907, xxiv. 13. See also "Microscope, Optics of the."

topf ultramicroscope, which is sometimes known as the "slit ultramicroscope." The optical system of the instrument is shown diagrammatically in *Fig.* 1. A beam of light from the sun or an arc lamp is focussed by means of an objective O_1 (focal length about 10 mm.) on a fine adjustable horizontal slit S. A reduced image of the slit is formed at F_1 by a second objective O_2 (focal length about 80 mm.). A narrow conical beam of light is formed at F_2 by a microscope objective M_1 acting as a condenser. If now a colloidal solution, such as a solution of colloidal gold particles, or a piece of ruby glass, which contains colloidal gold in suspension, be placed at F_2, the light scattered by the particles can be observed with a microscope M_2, whose optical axis is at right angles to the axis of the illuminating beam. What one observes is not, of course, the individual particles, but the diffraction discs formed by the light scattered from them, the general background being dark. In order to screen off any light that may be reflected from the edges of the slits an iris diaphragm I is placed between S and O_2. If desired, a polariser may be inserted between S and I. When an immersion objective is used in the observing microscope M_2, it is necessary to place a chisel-shaped diaphragm at D so as to cut off one-half of the illuminating beam and thus prevent objectionable reflection from the mount of the front lens, due to the closeness of the objective. In the slit ultramicroscope manufactured by Zeiss the optical parts of the illuminating system, together with the slit and diaphragms, are mounted on stands fitted to an optical bench.

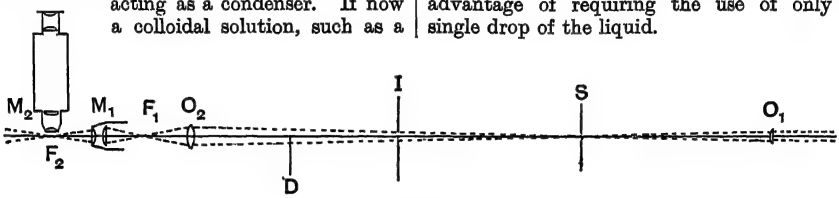


FIG. 1.

An improved form of this type of ultramicroscope was introduced some years ago by Zsigmondy,[1] who suggested as its designation the "immersion ultramicroscope." The salient feature of the new instrument is the employment of an immersion objective at M_1 in the illuminating system as well as in the observing system. In order to make this possible, portions of the front and meniscus lenses and of the mounts of the two immersion objectives are ground away at an angle of 45°, so that the objectives can be brought close enough together. The objectives have a focal length of 6 mm. and numerical aperture 1·05, the front lenses being constructed of fused quartz. The chief advantage of the new design is the greater obtainable intensity of the diffraction images, for the intensity under similar circumstances is proportional to the product of the numerical apertures of the illuminating and observing objectives. For the examination of colloidal solutions the instrument has the further advantage of requiring the use of only a single drop of the liquid.

(ii.) *Oblique Illumination*—(*a*) *Non-axial Beam without Central Stop.*—This method of illumination has been employed in different forms by microscopists for increasing the resolving power of an objective. It was first adapted for ultramicroscopic work by Cotton and Mouton;[2] the arrangement they used is illustrated diagrammatically in *Fig.* 2. A drop of the solution under examination is placed under a cover-glass on a microscope slide S, which is laid on a special block of glass G, optical contact between S and G being obtained by inserting a thin layer of cedar oil. An oblique beam of light is reflected internally from the base of G and then totally reflected at the cover-glass, so that no direct light enters the objective of the observing microscope M. The beam of light is brought to a focus in the drop of the solution by means of a condensing lens L, so that any small particles in the solution which come into the path of the beam are strongly illuminated and in consequence scatter light in all directions. Thus the particles appear as bright specks on a dark background, as in the previous case. A Fresnel rhomb may be used instead of the block G illustrated in the diagram.

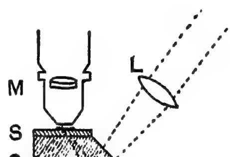


FIG. 2.

(*b*) *Axial Beam with Central Stop in Substage Condenser.*—This is another form of oblique illumination which has the advantage over the method (ii.) (*a*) of being more compact, the illuminating system forming a component part of the observing microscope. The principle of the method is illustrated in *Fig.* 3,

[1] R. Zsigmondy, *Phys. Zeits.*, 1913, xiv. 975.

[2] A. Cotton and H. Mouton, *Les Ultramicroscopes et l'objets ultramicroscopiques; Comptes Rendus*, 1903, cxxxvi. 1657; *Bull. Soc. Fr. Phys.*, 1903, p. 54.

which represents three types of sub-stage condenser specially adapted for dark-ground illumination. *Fig.* 3 (*a*) shows a vertical section of the Wenham paraboloid condenser as applied by Siedentopf.[1] It consists essentially of a truncated paraboloid of glass G, with the top of which a microscope slide S makes optical contact, a drop of the solution to be examined being placed under a cover-glass C. A central stop A below the condenser serves to cut out the central portion of the illuminating beam which is reflected upwards from a plane microscope mirror. The rays which enter the paraboloid are internally reflected at its surface and brought to a focus in the drop of solution, where they are totally reflected at almost grazing incidence. Only the light which is scattered by the particles in the solution can then enter the objective of the observing microscope. The paths of two rays X, Y of the illuminating beam are illustrated in the diagram, the directions being shown by means of arrow-heads.

A similar type of illuminating system is obtained with the cardioid condenser which Siedentopf[2] applied to the so-called "cardioid ultramicroscope." It depends on the principle that the image formation obtained by combining a spherical reflecting surface with a cardioid surface is aplanatic—that is, spherical aberration is eliminated, as is the case with the paraboloid condenser, and in addition the sine condition is fulfilled. Approximate aplanatism is attained if the cardioid surface be replaced by a spherical surface. The cardioid condenser is illustrated diagrammatically in *Fig.* 3 (*b*); it consists essentially of two glass portions G_1, G_2 cemented together, the surface of separation *a* being spherical. The rays of the illuminating beam are reflected outwards at the spherical surface *b*, and then inwards at the spherical surface *c*, coming to a focus at the object as in the previous case. An annular diaphragm is provided at B in addition to the one at A.

The Ignatowski condenser,[3] which resembles the cardioid system, is shown in *Fig.* 3 (*c*). It also consists of two glass portions G_1, G_2 cemented together. In all these forms of condenser the numerical aperture of the illuminating rays is of the order of 1·0 to 1·4.

(*c*) *Axial Illumination with Central Stop behind Objective.*—In this method the direct rays of the illuminating beam are stopped by a central diaphragm at the back of the microscope objective or by a stop formed by grinding flat and blackening a small central portion of the curved surface of the front lens of the objective. The method has the advantage of being free from centering troubles, but it suffers from the fact that the central stop alters the brightness distribution in the diffraction fringes.

§ (2) PHYSICAL APPLICATIONS OF THE ULTRAMICROSCOPE.—Although the chief applications of ultramicroscopy have been in the realms of colloidal chemistry, there are a number of important physical problems which have been developed as the result of ultramicroscopical research. For example, the study of the so-called Brownian movement has confirmed the fundamental hypotheses of the kinetic theory of liquids and gases, and contributed evidence of the existence of the molecule. The term "Brownian movement" derives its name from the work of the botanist Brown, who observed that small particles of solid matter, when suspended in liquid solution, exhibit rapid to-and-fro motions.[4] The complete investigation and explanation of this movement has been almost entirely due to the work that has been made possible by the invention of the ultramicroscope. When a colloidal solution, of gold for example, is examined by means of an ultramicroscope, the particles are seen to move about in all directions, the motion of the bright specks of light exhibiting a beautiful scintillating effect. It is impossible in the scope of this article to discuss all the details and consequences of the theories that

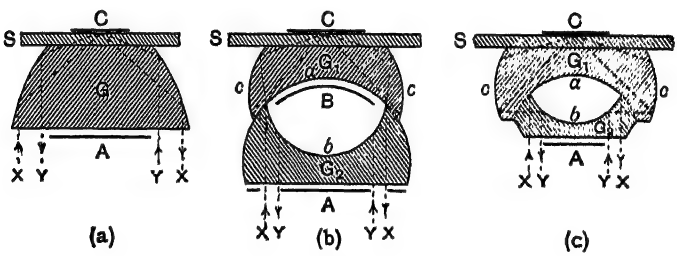


FIG. 3.

[1] H. Siedentopf, *Zeits. f. wiss. Mikroskopie*, 1907, xxiv. 104.
[2] H. Siedentopf, *Ber. d. D. Phys. Ges.*, 1910, xii. 6.
[3] W. v. Ignatowski, *Zeits. f. wiss. Mikroskopie*, 1908, xxiv. 26, 64; 1909, 387.
[4] R. Brown, *Phil. Mag.*, 1828, iv. 161.

have been put forward to explain the Brownian movement, but it may be stated that it has practically been proved, both theoretically and by observation, that it is mainly, if not entirely, due to the collisions with the molecules of the medium in which the particles are suspended.

Smoluchowski,[1] basing his calculations on the kinetic theory, and making use of Stokes' law, worked out theoretically that the mean distance D_x, traversed in time t by a small sphere of radius a, moving in a liquid whose coefficient of viscosity is η, is given by

$$D_x^2 = \tfrac{32}{27} \cdot \frac{c^2 mt}{3\pi\eta a} = \tfrac{32}{27} \cdot \frac{R}{N} \cdot \frac{T \cdot t}{3\pi\eta a},$$

where c = mean velocity of the molecules of the medium,
m = mass of each molecule of the medium,
R = gas constant,
T = absolute temperature,
N = number of molecules in one gram-molecular weight of any substance.

Einstein,[2] by applying the laws of osmotic pressure to the motion of small particles in a liquid medium, developed the formula

$$D_x^2 = \frac{R}{N} \cdot \frac{T \cdot t}{3\pi\eta a},$$

which differs from Smoluchowski's formula in not having the factor 32/27.

Langevin,[3] from a consideration of the equation of motion of a particle moving under the action of a force due to molecular shocks, obtained the same result as Einstein.

The results of the ultramicroscopic measurement of D_x by different observers, for a large variety of particles of different sizes, have been shown to agree very closely with the values obtained from the formulae thus developed. Furthermore, observations by Perrin [4] on the time of rotation of comparatively large grains of mastic have been found to fit in with the formula, developed by Einstein, for the mean square of the angle of rotation d, in time t, namely,

$$d^2 = \frac{R \cdot T}{N} \cdot \frac{t}{4\pi\eta a^3}.$$

Perrin [5] also made observations on the relative distribution of colloidal particles of gamboge and mastic at different heights in vertical columns of their solutions, the results again agreeing with the theoretical values obtained from the kinetic theory. The remarkable agreement that has been found between observation and theory may be looked upon as a proof of the reality of the molecular and kinetic hypotheses.

A number of interesting experiments have been made in connection with the motion of ultramicroscopic particles (smoke, silver, gold, mercury, etc.) in gases.[6] Such particles have been found to be electrically charged, and by measuring the rates with which they fall in the field of gravity and in an electrical field of known intensity it has been possible to determine the charges carried by the particles. It has been found that the charge on each particle is the elementary electrical charge (that is, the charge of an electron) or a multiple of this quantity, the values obtained agreeing very closely with those derived from radioactive and electronic measurements.

Another important application of the ultramicroscope to physical problems is the study of the optical effects exhibited by colloidal solutions. The late Lord Rayleigh [7] has shown theoretically that the intensity I of the light scattered by particles, whose size is small in comparison with the wave-length of the light employed, is given by

$$I \propto I_0 \frac{(D^1 - D)^2}{D^2}(1 + \cos^2\beta)\frac{m\pi T^2}{\lambda^4},$$

where I_0 = intensity of incident light,
D' = optical density of particles (proportional to square of refractive index),
D = optical density of medium,
m = number of particles per unit volume,
T = volume of a particle,
λ = wave-length of scattered light,
β = angle between line of sight and direction of incident light.

It follows that for a given distribution of small particles the percentage of violet light scattered is very much greater than that of red light.

The ultramicroscopic examination of the light scattered by the particles in colloidal solutions has confirmed the validity of Rayleigh's law. The scattered light is, in general, partially polarised, and often completely polarised in a certain plane. The whole problem is a very complex one and has engaged the attention of a large number of workers. One of the most important contributions to the subject is that of Garnett,[8] who investigated, both theoretically and experimentally, the absorption of various colloidal solutions, metal glasses, and metallic films, his mathematical treatment being based on the electromagnetic theory. His experimental results were found to be in good agreement with theory. The

[1] M. Smoluchowski, *Bull. Intern. Acad. d. Sci. Cracovie*, 1906, vii. 577; *Ann. d. Physik*, 1906 (4), xxi. 756.
[2] A. Einstein, *Ann. d. Physik*, 1905, xvii. 549; 1906, xix. 371.
[3] P. Langevin, *Comptes Rendus*, 1908, cxlvi. 530.
[4] J. Perrin, *Comptes Rendus*, 1909, cxlix. 549.
[5] *Ibid.*, 1908, cxlvii. 530, 594; *Zeits. f. Elektrochemie*, 1909, xv. 269.
[6] Cf. M. de Broglie, *Comptes Rendus*, 1908, cxlvi. 624, 1010; 1909, cxlviii. 1163, 1315; F. Ehrenhaft, *Wien. Sitzber. Natur-Wissen.*, 1907, cxvi. 1139; *Phys. Zeits.*, 1914, xv. 952; 1915, xvi. 227; R. A. Millikan, *Phil. Mag.*, 1910 (6), xix. 209; *Phys. Rev.*, 1911, xxxii. 350; 1913, ii. 2nd Ser. 109; *Phys. Zeits.*, 1913, xiv. 796.
[7] Lord Rayleigh, *Phil. Mag.*, 1871 (4), xli. 107, 447; 1899 (5), xlvii. 375.
[8] J. C. M. Garnett, *Roy. Soc. Phil. Trans.* (A), 1904, cciii. 385; 1906, ccv. 237.

problem of the scattering of light by small particles is very important in connection with the theory of the colour of the sky.

From the above brief sketch of some of the problems involved in the examination of ultramicroscopical solutions it may be seen that the invention of the ultramicroscope has opened up a wide field of physical research which has already yielded important results in connection with molecular phenomena.

J. S. A.

Uniaxial Crystal: a crystal having one optic axis. See "Polarised Light and its Applications," §§ (5), (6), and (18).

Unit Points and Planes of a Lens. See "Objectives, Testing of Compound," § (1); also "Lens Systems, Aberrations of."

Uranium X: the first product of the radioactive disintegration of uranium, discovery and properties of. See "Radioactivity," § (2).

— V —

Valves of Brass Wind-instruments, Faulty Intonation of Ordinary. See "Sound," § (39).

Valves on Cornet, tabulated. See "Sound," § (39), Table IX.

Variable Power Eyepieces. See "Eyepieces," § (9).

Vector Method of deducing Properties of Diffraction Spectra. See "Diffraction Gratings, Theory," § (5).

Vernier. See "Divided Circles," § (11).

Vernon-Harcourt Pentane Lamp: a flame standard of light of 10 candle-power. See "Photometry and Illumination," § (6).

VIBRATIONS OF AIR IN A TUBE

In Acoustics and some other problems it is necessary to investigate the small vibrations of a compressible fluid such as air in a cylindrical tube. We assume that the motion of each particle is parallel to the axis of the tube, so that particles which were originally in a plane at right angles to the axis remain always in such a plane.

Let ξ be the displacement at time t of the particles originally in a plane at a distance x from the origin. Let p_0, ρ_0 be the undisturbed pressure and density of the air, p, ρ the values of these quantities at time t. The matter which was originally between two planes at distances x and dx from the origin lies at time t between planes at distances $x+\xi$ and $x+\xi+(1+d\xi/dx)dx$. Its density was ρ_0, it is now ρ; hence if A be the area of the cross-section of the tube,

$$A\rho_0 dx = A\rho\left(1+\frac{d\xi}{dx}\right)dx,$$

$$\therefore \rho = \frac{\rho_0}{\{1+(d\xi/dx)\}}. \qquad (1)$$

The force acting on this matter on the face near the origin is Ap, while on the opposite face it is $-A\{p+(dp/dx)dx\}$. Thus the equation of motion is

$$A\rho_0 dx\frac{d^2\xi}{dt^2} = -A\frac{dp}{dx}dx$$

or

$$\rho_0\frac{d^2\xi}{dt^2} = -\frac{dp}{dx}. \qquad (2)$$

To complete the solution we need to know the relation between p and ρ.

If the temperature is constant, we have

$$p = \frac{p_0}{\rho_0}\cdot\rho. \qquad (3)$$

If, as is the case with sound waves, the compressions and rarefactions of the air take place so rapidly that there is no loss or gain of heat, we have [1]

$$p = k\rho^\gamma = p_0\left(\frac{\rho}{\rho_0}\right)^\gamma. \qquad (4)$$

Combining (1) and (4) then we get

$$p = \frac{p_0}{\{1+(d\xi/dx)\}^\gamma} \qquad (5)$$

and

$$\frac{dp}{dx} = -\frac{\gamma p_0}{\{1+(d\xi/dx)\}^{\gamma+1}}\frac{d^2\xi}{dx^2}.$$

Thus the equation of motion is

$$\rho_0\frac{d^2\xi}{dt^2} = \frac{\gamma p_0}{\{1+(d\xi/dx)\}^{\gamma+1}}\frac{d^2\xi}{dx^2}, \qquad (6)$$

and remembering that we are dealing only with small motions we may neglect $d\xi/dx$ in the denominator and obtain

$$\frac{d^2\xi}{dt^2} = \frac{\gamma p_0}{\rho_0}\frac{d\xi}{dx^2}. \qquad (7)$$

Had we assumed isothermal conditions we should clearly have found

$$\frac{d^2\xi}{dt^2} = \frac{p_0}{\rho_0}\frac{d^2\xi}{dx^2}. \qquad (8)$$

These can be put into a slightly different form thus: The elasticity E of a gas is

[1] See article "Thermodynamics," § (15), Vol. I.

measured by the ratio of the small change of pressure required to produce a change of volume to the change per unit volume; thus if $-dv$ is the decrement of a volume v produced by an increment of pressure dp,

$$E = -v\frac{dp}{dv}. \qquad . \quad . \quad . \quad (9)$$

Let M be the mass, ρ the density of the volume v, v_0 and ρ_0 being initial values.

Then $\qquad v\rho = M = v_0\rho_0$

or $\qquad v = \frac{v_0\rho_0}{\rho};$

hence $\qquad dv = -\frac{v_0\rho_0}{\rho^2}d\rho$

and

$$E = -\frac{v_0\rho_0}{\rho}\frac{dp}{dv} = \rho\frac{dp}{d\rho}$$

$$= \rho k\gamma\rho^{\gamma-1}$$

$$= k\gamma \, . \, \rho^\gamma = \gamma p,$$

since $\qquad p = k\rho^\gamma.$

Or, denoting by E_0 the elasticity in the undisturbed state, we have

$$E_0 = \gamma p_0; \qquad . \quad . \quad . \quad (10)$$

hence

$$\frac{d^2\xi}{dt^2} = \left(\frac{E_0}{\rho_0}\right)\frac{d^2\xi}{dx^2}.$$

If we put $c^2 = E_0/\rho_0$, this takes the form

$$\frac{d^2\xi}{dt^2} = c^2\frac{d^2\xi}{dx^2}, \qquad . \quad . \quad . \quad (11)$$

and is the same equation as is discussed in the article "Strings, Vibration of." The general solution is

$$\xi = f(ct-x) + F(ct+x), \quad . \quad . \quad (12)$$

as can readily be verified, and represents, as there shown, waves travelling in the positive and negative directions with the velocity c or $\sqrt{(E_0/\rho_0)}$, *i.e.* the square root of the ratio of the elasticity to the density.

The statement in this form can be extended to other cases of sound motion, the term elasticity being defined with regard to the circumstances of each case.[1]

VIBRATIONS OF A ROD

IMAGINE a straight rod of uniform section A in a state of motion in which every part is moving parallel to the axis so that all points in any plane at right angles to the axis always remain in such a plane. Let ξ be the displacement at time t of a plane initially at a distance x from one end of the rod, and consider the motion of the particles in the section between this plane and one initially at a distance $x+dx$; they will at time t lie between the planes $x+\xi$ and $x+\xi+(1+(d\xi/dx))dx$.

The extension or "stretch" of the short length dx is $d\xi/dx$, and if q represents Young's modulus for the rod (see article "Elasticity," § (4), Vol. I.), the force on the face of the section nearest the origin is $-qA(d\xi/dx)$, that on the opposite face is $qA(d\xi/dx)+(d/dx)(qA(d\xi/dx))dx$. Thus the resultant force on the section is

$$qA\frac{d^2\xi}{dx^2}dx.$$

Hence, if ρ be the density, the equation of motion of the section is

$$\rho A\frac{d^2\xi}{dt^2}dx = qA\frac{d^2\xi}{dx^2}dx$$

or

$$\frac{d^2\xi}{dt^2} = \frac{q}{\rho}\frac{d^2\xi}{dx^2}. \qquad . \quad . \quad . \quad (1)$$

If we put $q/\rho = c^2$, this, it can be shown, is satisfied by

$$\xi = f(ct-x) + F(ct+x), \quad . \quad . \quad (2)$$

and represents[2] wave motion travelling with velocity c. The value of c is $\sqrt{q/\rho}$, or, as in other cases, the speed of the waves is determined by the square root of the ratio of the elasticity to the density, the elasticity in this case being measured by Young's modulus.[3]

VIERORDT SPECTROPHOTOMETER. See "Spectrophotometry," § (12).

VIOLIN FAMILY OF STRINGED INSTRUMENTS. See "Sound," § (30).

VIOLLE STANDARD: an incandescence standard of light. See "Photometry and Illumination," § (9).

VISCOSITY OF GLASS. See "Glass," § (33).

VISIBILITY. See "Eye," § (6).

VISUAL METHODS OF OBSERVATION, used in microscopy with ultra-violet light, of animal tissues and other substances which fluoresce. See "Microscopy with Ultra-violet Light," § (7).

VITREOUS HUMOUR: one of the fluids contained in the eyeball. See "Eye," § (2).

VOICE, HUMAN. See "Sound," § (45).

Compared with other musical instruments. See *ibid.* § (46).

[1] See articles "Vibrations of a Rod" and "Sound" § (12).

[2] See article "Strings, Vibrations of."

[3] See articles "Sound" § (12), and "Vibrations of Air in a Tube."

— W —

WATCH-POCKET SPHEROMETER. See "Spherometry," § (6).

WATER, ACTION ON GLASS. See "Glass, Chemical Decomposition of," § (1).

WAVE FORM OF POTENTIAL EXCITING X-RAYS. See "Radiology," § (26).

WAVE-LENGTH, VARIATIONS OF, due to—(1) the pressure of gas in the luminous source; (2) the pole effect; (3) the presence of isotopes; (4) the presence of a strong electric field. See "Spectroscopy, Modern," § (8).

WAVE-LENGTH AND REFRACTIVE INDEX. See "Optical Glass," § (3).

WAVE-LENGTHS, THE MEASUREMENT OF

§ (1) INTRODUCTION. — The phenomena of coloured rings, commonly known as "Newton's Rings," were described in his *Opticks*, published in 1704, but were not explained on the basis of the undulatory theory of light until 1802, when Thomas Young (1) put forth the theory of interference of light waves, and was thereby led to make the first determinations of wave-lengths of light corresponding to the different colours. Measurements of wave-lengths corresponding to narrow spectral lines were not made until somewhat later, after the discovery of the dark lines in the solar spectrum. Newton showed by means of a glass prism that sunlight was made up of various colours, and to all observers the solar spectrum appeared to be a continuous graduation of colours until Wollaston (2) in 1802 observed distinct dark lines, which he regarded as boundaries to intervals of colour. Wollaston's discovery was forgotten until 1814, when Fraunhofer (3) repeated it and recognised that the dark lines represented definite wave-lengths in the solar spectrum. In his investigations, Fraunhofer mapped about 700 of these lines, assigning letters of the alphabet to the most prominent ones, beginning with A in the red and ending with H in the violet. These lines, still called the Fraunhofer lines, are familiar to every one who has any knowledge whatever of spectroscopy. Investigation into the possibilities of measuring the actual lengths of light waves led Fraunhofer to develop the theory of interference of light and invent the first diffraction gratings, with which he made wave-length measurements of considerable accuracy. His method (4) for measuring the wave-lengths of light by means of a grating is indeed the same as that of the present day. It is based upon the interference of rays which are diffracted upon passing through a set of equal and equidistant apertures or slits constituting the grating.

It was not until 1859 that the physical significance of the Fraunhofer lines became known. Kirchhoff and Bunsen then explained that "the dark lines of the solar spectrum, which are not caused by the terrestrial atmosphere, arise from the presence in the glowing solar atmosphere of those substances which in a flame produce bright lines in the same position." Inasmuch as these two scientists proved the existence of many terrestrial elements in the sun, it was a natural consequence that the solar spectrum itself should become the standard of reference for wave-lengths of light.

With the aid of gratings ruled on glass, Ångström made, as accurately as possible, a map of the Normal Solar Spectrum, which he published in 1868. The wave-length measurements covered the visible spectrum between the Fraunhofer lines A and H, and were expressed in ten-millionths of a millimetre to 2 decimal places. This unit of length, 1×10^{-10} metre, sometimes called the "tenth metre," but more commonly known as the "Ångström Unit" (symbolised simply by A), is now almost universally used in measurements of wave-lengths of light.

Ångström's map was superseded by Rowland's Normal Map of the Solar Spectrum, published in 1888. and based upon his invention of the concave grating. This photographic map was about 20 metres long, the maximum error in any part of the scale of wave-lengths being estimated as less than 0·01 A. Rowland's Preliminary Table of Solar Spectrum Wave-lengths, representing over 20,000 lines, was based upon determinations of the absolute wave-length of one of the D lines of sodium. The wave-lengths were given to 3 decimals in A, Rowland believing that the absolute values were correct to 1 part in 100,000, and that the errors in relative value were not greater than a millionth. For a period of twenty years practically all spectroscopic measurements were based on these values, and even at the present time they are widely used.

§ (2) THEORY AND USE OF DIFFRACTION GRATINGS IN WAVE-LENGTH MEASUREMENTS. —The simple theory[1] of the plane diffraction grating gives the wave-length of light as a function of the grating space and of the angles of incidence and diffraction (or reflection) of the light rays. *Fig.* 1 represents the cross-section of an element of a transmission grating

[1] See "Diffraction Gratings, Theory of"; also "Diffraction Gratings, Manufacture and Testing of."

in which AB′ is to be regarded as the grating space b. If light from an illuminated slit S falls as a parallel beam on the grating at an angle of incidence i with the normal, a large portion passes directly through the grating apertures, but a part of it is diffracted in the direction AC, making the angle θ with the normal. By drawing the perpendiculars BB′ and CB′, it is seen that the difference in path travelled by corresponding rays from adjacent apertures, or the retardation as it is called, is given by the sum of the lengths BA and AC, and therefore when this path difference is equal to some integral number of whole wave-lengths $n\lambda$ bright images of S in colour

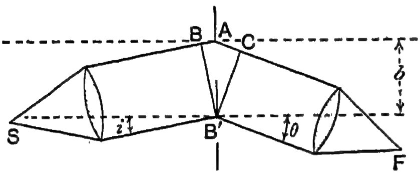


Fig. 1.

corresponding to λ are produced when the diffracted rays are brought to a focus. This condition is easily seen to be given by $BA + AC = n\lambda = b(\sin i + \sin \theta_n)$, in which n represents the order number of the spectrum.

Reflection gratings have been more successfully made than transmission gratings, and have, therefore, been more extensively used for wave-length measurements, but the law of the gratings ruled on polished plane metal surfaces is quite similar to that of gratings ruled on glass. A beam of light falls on the grating G, *Fig.* 2, making an angle of incidence i with the normal, and part of the light is diffracted at an angle θ. The perpendiculars BB′ and CC′ are drawn to show that B′C and BC′ are the retardations. In this case the total retardation is seen to be determined by $B'C - BC' = b(\sin i - \sin \theta)$. Wherever this equals one or more whole wave-lengths, $n\lambda$, a bright image is seen in colour corresponding to λ, and we have the law of the grating

$$n\lambda = b(\sin i - \sin \theta_n).$$

If the diffracted rays are on the same side of the normal as the incident rays, then the two retardations are added, so that the general equation for all the spectra must be written

$$n\lambda = b(\sin i \pm \sin \theta_n),$$

the positive or negative sign being used when the incident and diffracted rays are on the same or opposite sides of the normal respectively.

Depending on the position of the grating in relation to the collimating and observing telescope, there are five different methods of making wave-length measurements with plane gratings. To sketch these the wave-length equation or grating law may be written

$$n\lambda = b\{\sin i + \sin(\delta - i)\},$$

in which δ is the angle of deviation or $\delta = i + \theta$. (1) The plane of the grating is placed perpendicular to the collimator, so that $i = 0$ and $n\lambda = b \sin \delta$. This method was used by Fraunhofer with the first diffraction gratings for the first absolute measurements on wave-lengths of light. (2) The grating is normal to the telescope, so that $\theta = 0$ and $n\lambda = b \sin \delta$ as before. (3) Both i and θ are given values, so that $n\lambda = b(\sin i + \sin \theta)$. This was the method adopted by Ångström. (4) The deviation δ is a minimum. Writing the above equation in the form

$$n\lambda = 2b \sin \frac{(i+\theta)}{2} \cos \frac{(i-\theta)}{2},$$

it is seen that $\cos(i-\theta)/2$ is a maximum when $i = \theta$, which is the condition for minimum

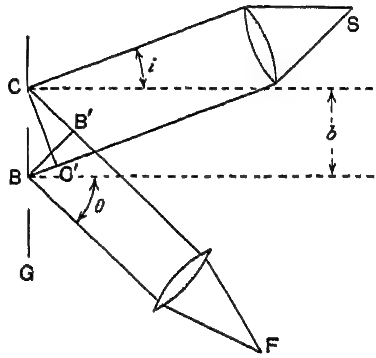


Fig. 2.

deviation δ. Then the wave-length is found from

$$n\lambda = 2b \sin \frac{\delta}{2}.$$

(5) The telescope and collimator are clamped at an angle α, and the grating rotated through the measured angle β, whence it can be shown that

$$n\lambda = 2b \sin \beta \cos \frac{\alpha}{2}.$$

In this case observations are made by adjusting the required spectral line upon the cross-wire in the eyepiece, and then rotating the grating until the reflected image is brought upon the cross-wire.

All of these methods require the measurement of the grating space b; (1), (2), and (4) require the measurement of one angle; while

(3) and (5) require the measurement of two angles.

Bell (5), whose determinations of the wave-length corresponding to D_1 received greatest weight in establishing Rowland's standards, used the second method with two glass transmission gratings and the fifth with two reflection gratings ruled on speculum metal.

All plane gratings require the use of lenses to focus the diffracted waves. These introduce chromatic difficulties in focussing, and further limitations due to absorption, so Rowland, in 1881, invented gratings ruled on spherical mirrors, so that the concave grating should focus spectra without the use of a lens. The most important characteristic of the concave grating is the simplicity with which it can be focussed. If the slit of the spectroscope and the grating be placed on the circumference of a circle[1] whose diameter equals the radius of curvature of the grating, the spectra are all in focus on this circle, illustrated by *Fig.* 3,

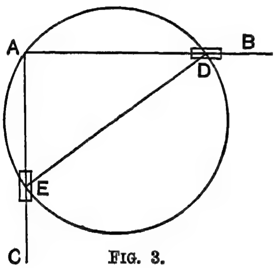


FIG. 3.

DE being the diameter of the circle and equal to the radius of curvature of the concave grating. If the slit be placed on the circumference of this circle, *e.g.* at A, the spectra are all in focus on the circumference, as at E.

With both plane and concave gratings there is great advantage obtained by forming the spectra on the normal to the reflecting surface, for this gives the so-called normal spectrum, that is, distances measured either way from the grating normal are proportional to the change in wave-length observed. Rowland devised a mounting for his concave grating which automatically kept the observing apparatus (eyepiece or camera) on the normal to the grating surface. The grating and camera, *Fig.* 3, are mounted on carriages rigidly separated by the beam DE, whose ends are constrained to move along the rails AB and AC. These rails are adjusted at right angles to each other, and have the slit at their intersection A. Every circle having the shifting beam DE as a diameter must pass through A, and if the grating is fixed normal to DE, its centre of curvature coincides exactly with E, giving a normal spectrum at that point.

[1] See "Diffraction Gratings, Theory of," § (11), Vol. IV.

Rowland's method of determining relative wave-lengths is based on a simple feature of this grating mounting in that the wave-length equation is reduced to $n\lambda = b \sin i$, when observations are made near the perpendicular to the grating. Then for a given angle i and different spectral orders, 1, 2, 3, etc.,

$\lambda_1 = b \sin i$ in the spectrum of the first order,
$2\lambda_2 = b \sin i$ in the spectrum of the second order,
$3\lambda_3 = b \sin i$ in the spectrum of the third order,
etc.,

where λ_1, λ_2, λ_3, etc., are the wave-lengths in the 1st, 2nd, 3rd, etc., orders. It follows that $\lambda_1 = 2\lambda_2 = 3\lambda_3$, etc., which means that various orders of spectra are superposed, and the wave-lengths of each are inversely proportional to the order number of the spectrum. For example, a wave-length of 6000 A in the first order spectrum will have superposed on it 3000 A in the second order and 2000 A in the third order. By measuring the lines in various orders nearly superposed on the sodium line (D_1 = 5896·156 A), Rowland first compared 14 solar wave-lengths with this standard as accurately as possible, and in a similar way obtained from these the wave-lengths of principal lines throughout the solar spectrum (6).

Another type of concave grating mounting which has been adopted by many laboratories is due to Runge and Paschen (7). This uses a lens or concave mirror to illuminate the grating with parallel light from the slit, and reduces the dispersion or scale of spectrum to about half that in the Rowland mounting, with the result that the intensity of the spectrum is theoretically quadrupled. The astigmatism in spectra given by the Rowland mounting is practically eliminated by mounting the grating in parallel light. These stigmatic slit images permit (1) the use of diaphragms in front of the slit to photograph different spectra in juxtaposition; (2) the projection of different parts of sources, of interferometer fringes, etc., on the slit for spectral analysis.

The complete theory of diffraction gratings, both plane and concave, together with detailed directions on mounting, adjusting, and using gratings for wave-length measurements, may be found in vol. i. *Handbuch der Spectroscopie*, by H. Kayser, or in *Spectroscopy*, by E. C. C. Baly. Limitations of space will not permit more on this subject here, except to call attention to a recent controversy on the validity of the fundamental law of the grating, and questioning the use of gratings for accurate measurements of wave-lengths. When in 1901 Rowland's standards were found to contain relative errors of nearly 1 part in 100,000, these errors were attributed to the erratic behaviour of diffraction gratings.

Kayser (8), in 1904, made some experiments which led him to the conclusion that the coincidence between orders of spectra is not to be depended upon. It has since been shown (9), however, that the fundamental law of diffraction gratings is true, so that the errors in Rowland's values as well as the experience of Kayser require other explanations. According to Jewell (10), the relative errors in Rowland's Table are due to disturbed apparatus and failure to recognise that corrections were necessary for the Doppler-Fizeau effect or for air-pressures and temperatures, while Goos (11) has shown that the discrepancies observed by Kayser may be explained by slight defects in the adjustment of the grating and perhaps to inattention to constant conditions in operating the sources of light. If accurate measurements of wave-lengths are to be made with gratings, stability of mounting and correctness in adjustment must be insured and temperature fluctuations guarded against. The change in the refractive index and dispersion of air (12) will cause overlapping spectrum orders observed as coincidences at one temperature to be shifted at other temperatures.

Another matter which deserves attention in connection with the use of gratings for wave-length observations is the presence of spurious lines or "ghosts," due to periodic errors in the ruling. These false lines are present in all grating spectra, and cannot be distinguished from real lines except by testing their relationship to parent lines by means of numerical ratios. Two general types of such spurious lines are known: (1) Ghosts located near the spectral line, and symmetrically placed on both sides of it; and (2) spurious lines, whose positions in the spectrum are far removed from the parent line.

False lines of the first type, generally known as the Rowland ghosts, arise from errors of ruling repeated with successive revolutions of the screw. For example, Rowland gratings ruled 20,000 lines per inch with a screw of 1/20 inch pitch have an interval of 1000 lines in which errors recur. The position of these ghosts with respect to parent lines is that of additional spectra from a grating having 1000 times the grating space; their apparent wave-lengths differ from that of the parent line by $n/1000$ times the wave-length of the corresponding line if n represents the order of the false spectrum. Anderson ruled gratings with 750 lines per revolution of a screw having 1/20 inch pitch, thus making 15,000 lines per inch, while the use of every other or every third tooth of the wheel produced gratings of 7500 and 5000 lines per inch. These gratings give ghosts at distances $n\lambda/750$, $n\lambda/375$, and $n\lambda/250$ respectively, from the parent line of wave-length λ.

Except in complicated spectra, the Rowland ghosts are usually easy to detect on account of their symmetry and proximity to the real lines, but it is much more difficult to eliminate spurious lines of the second type. Lyman (13) found spurious doublets occurring at positions corresponding roughly to $\lambda/3$ and $2\lambda/3$ in spectra produced by Rowland gratings having 14,438 lines per inch. An Anderson grating with 7500 lines per inch has been found at the Bureau of Standards to give false reproductions of a given line of wave-length λ in positions near $2\lambda/5$, $3\lambda/5$, $4\lambda/5$, $6\lambda/5$, $7\lambda/5$, $8\lambda/5$, and $9\lambda/5$. Similar false spectra have been detected in the 5000 lines per inch gratings ruled by using every third tooth in the wheel. Although these secondary lines are usually of relatively low intensity, they may, nevertheless, be very troublesome and embarrassing when recorded by the long exposures necessary to photograph the infra-red and extreme ultra-violet spectra.

§ (3) Standard Wave-lengths by Interferometer Methods.—An extraordinary degree of accuracy in wave-length measurements was made possible with the invention of the interferometer apparatus, in which a beam of light coming from a single source is first separated into two portions which are retarded relative to each other by passing over unequal optical paths, and then reunited, producing interference fringes. The interferometer method of wave-length measurement has several advantages over the grating method: the distance between two parallel interferometer plates is easily and accurately determined; it eliminates the precise measurement of angles; it permits the production of various lengths which are exact multiples of some unit; and less labour is required to make accurate comparisons of wave-lengths.

The first absolute measurement[1] of a light wave by the interferometer method was made in 1893 by Michelson (14), who determined the number of waves of three radiations from cadmium vapour which were equal in length

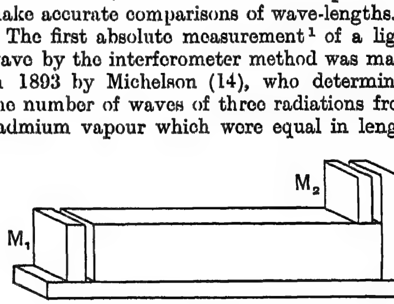


Fig. 4.

to the standard metre. Nine intermediate standards, the lengths of which were represented by the distances between the planes of two mirrors M_1 and M_2, *Fig.* 4, were constructed. The lengths of these intermediate standards were 10 cm., 10×2^{-1} cm., 10×2^{-2} cm. . . . the smallest being 10×2^{-8}

[1] See also "Line Standards," § (7), Vol. II.

cm., or about 0·39 millimetre. The optical system and two of the standards A and B are shown in the diagram, *Fig.* 5, in which C is the mirror of reference. An observer at O

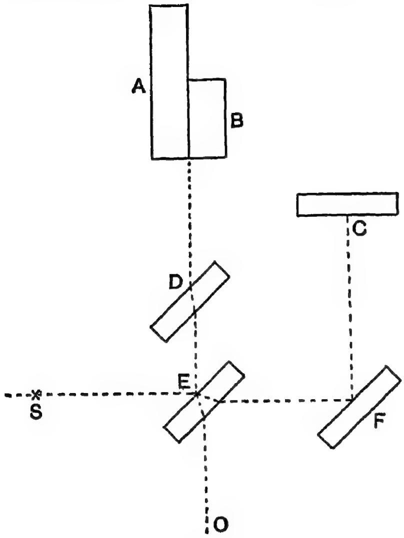


FIG. 5.

sees by reflection at the half-silvered surface E, the image of C, superposed on the four mirrors on A and B, two of which lie side by side below, and another pair above. B and C are mounted on carriages which can be moved back or forth by means of parallel screws. The length of the shortest standard was first determined in terms of the wave-length of cadmium red light. The front surface of B was adjusted to the same optical distance as C, and the reference plane C was then slowly light as a source and turning the mirrors slightly about a vertical axis gave vertical fringes, and the dark central fringe which marks zero difference of path appeared on adjacent edges of the front mirror of A and B. Then C was moved back until the dark fringes were seen in the upper mirror of B. Now when C was moved back through its own length the dark fringes again appeared in the lower mirror of B, and moving C back a distance equal to the length of B brought the fringes once more into view on the upper mirror of B. If A was exactly twice the length of B, fringes were also seen in the upper mirror on A. If A was not exactly twice B the central fringe was displaced, and the difference, in wave-lengths, between A and 2B was easily obtained from a slight tilt of the compensator D. In this way the lengths of successively larger intermediate standards were obtained in terms of wave-lengths of light. Finally the 10 cm. standard was compared with the metre by moving it through its own length 10 times. Michelson found at 15° C. and 760 mm. pressure

1 m. = 1553163·5 red waves from which λ = 6438·4722 A.
,, = 1900249·7 green ,, ,, ,, ,, = 5085·8240 A.
,, = 2083372·1 blue ,, ,, ,, ,, = 4799·9107 A.

An absolute accuracy of 1 part in 2,000,000 was claimed, while the relative accuracy might be 1 in 20,000,000.

An independent measurement of the wave-length of cadmium red radiation, in terms of the metre, was made in 1907 by Messrs. Benoit, Fabry, and Perot (15). Compared with Michelson's, their apparatus was simplified and arranged so as to reduce the dangers from

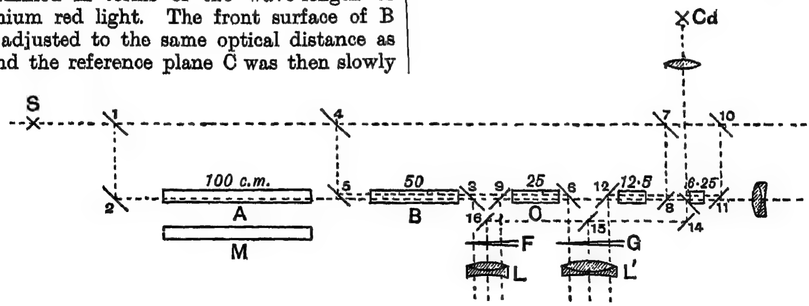


FIG. 6.

moved back to coincidence with the second surface of B, during which shift the number of circular fringes was counted as they flowed out from the centre of the interference figure.

The next operation was to compare twice the length of this standard B with the next longer one A. These were first adjusted so that their front lower mirrors were the same optical distance from E as C. Using white temperature changes in the standards by shortening the time necessary to perform the experiment. Five standards were made by separating half-silvered glass plates at distances of 6·25, 12·5, 25, 50, and 100 cm., and these were so placed that the comparisons could be made without moving any of them. The arrangement is shown diagrammatically in *Fig.* 6. As in Michelson's method, the

difference in length between one standard and the double length of the previous one had to be measured accurately. To compare A with B, for example, white light coming from S was reflected by mirrors 1 and 2, passed through A and B, mirror 5 being moved out of the way. This light was reflected by mirror 3 through a compensating wedge of air between two silvered glass plates F and into a microscope L. If A was very nearly 2B, a ray of light traversing A and twice reflected in B covered a path nearly equal to that of another ray crossing A twice and passing through B. Interference bands resulted, and from the position of the central fringe on the wedge compensator the difference between A and 2B was accurately determined in wave-lengths. Similarly, the other standards were compared, and finally the 6·25 cm. standard was measured in wave-lengths of the cadmium red radiation. The number of these waves in 1 metre was found to be 1,553,164·13, or $\lambda = 6438{\cdot}4696$ A in dry air at 15° C. and 760 mm. pressure. The probable error was one part in 10,000,000; this absolute determination of wave-length having the same accuracy as it is possible to obtain in the comparison of two metre bars. Their value differs from Michelson's by 1 part in 2,500,000, because it refers to dry air. An approximate reduction of Michelson's value to dry air gives $\lambda = 6438{\cdot}4700$ A, which checks the later value to 1 part in 16,000,000. The original memoirs should be referred to for further details concerning these beautiful experiments, which establish with extreme precision a wave-length of light as a unit of length probably more invariable, reproducible, and important than any other unit.

In 1907 the International Union for Co-operation in Solar Research made certain recommendations for the final establishment of a system of satisfactory light wave standards. The absolute value of the red radiation of cadmium as redetermined by Benoit, Fabry, and Perot was chosen as a primary or fundamental spectroscopic standard. It was further recommended that secondary standards be determined at intervals of about 50 A throughout the spectrum of the iron arc by comparison with the primary standard, using the interferometer method as devised by Fabry and Perot (16). This method is as follows: The red light from a cadmium lamp and the light from an iron arc are sent simultaneously through a Fabry and Perot interferometer, I, and the circular fringes thus formed are focussed with their centre on the slit S of a spectroscope, *Fig.* 7. The prism (or grating) separates the numerous spectral lines and permits, at P, an independent determination of the diameters of the circular fringe system produced by each line or radiation. The order of interference at the centre of a system of rings is determined for the two radiations whose wave-lengths are to be compared. Let λ be the wave-length of the cadmium red ray, and let P be the order of interference producing the first ring from the centre. The order at the centre is then $p = P + p'$, where p' is the fractional order, lying between zero and 1, to be determined from the diameter of the ring P. The whole number P is readily found from the distance between the plates. This results from trying various values of this distance with three or four wave-lengths until the computed fractional order of each agrees with its measured value (17). The order of interference at the centre ($i = 0$) is $p = 2e/\lambda$, where e is the distance between the plates, *Fig.* 8. In a direction making an angle i with the normal it is $2e \cos i/\lambda = p \cos i$.

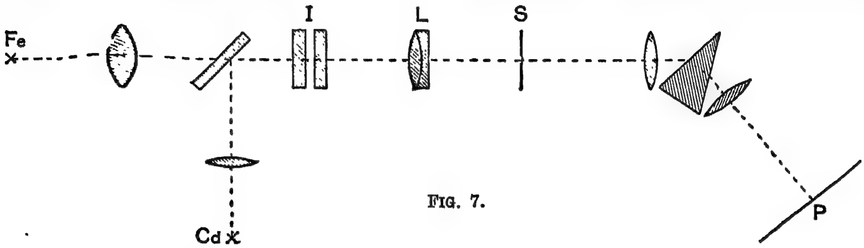


FIG. 7.

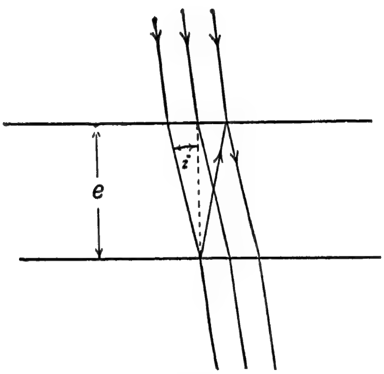


FIG. 8.

Now if x is the angular diameter of the ring P,

$$i=\frac{x}{2} \text{ and } P=p\cos\frac{x}{2}.$$

But
$$\cos\frac{x}{2}=1-\frac{i}{2}=1-\frac{x^2}{8}$$

approximately, so that

$$P=p\left(1-\frac{x^2}{8}\right), \text{ or } p=P\left(1+\frac{x^2}{8}\right)$$

and
$$p'=p-P=P\frac{x^2}{8}.$$

Similarly with a radiation of wave-length λ_1 we have

$$p_1'=P_1\frac{x_1^2}{8}.$$

The thickness of the air layer between the plates is assumed to be the same for all radiations, whence

$$\lambda(P+p')=\lambda_1(P_1+p'_1)$$

or
$$\lambda_1=\lambda\frac{(P+p')}{(P_1+p'_1)}, \quad \left(=\lambda\frac{p}{p_1}\right).$$

Substituting the above values of p' and p'_1,

$$\lambda_1=\lambda\frac{P}{P_1}\left(1+\frac{x^2}{8}-\frac{x_1^2}{8}\right).$$

If d represents the linear diameter of the ring and R the focal length of the lens which focusses the rings on the slit of the spectroscope, $x=d/R$ and

$$\lambda_1=\frac{\lambda P}{P_1}\left(1+\frac{d^2}{8R^2}-\frac{d_1^2}{8R^2}\right).$$

Slight corrections are necessary for the dispersion of the atmosphere and refractive index of air (18) at temperatures other than 15° C. and 760 mm., and for the reflection phase-change (19), which is a function of the wave-length. This method of wave-length comparisons can be applied only to waves whose lengths are already known accurately enough to allow P, the order of interference producing the first ring, to be determined without ambiguity. For this purpose an accuracy of one part in 20,000 or 30,000 is sufficient if the preliminary measurements are made with an air-plate thickness not exceeding 5 mm. The final accuracy of the comparison can be made to reach 1 part in 5,000,000.

§ (4) Standards of Wave-length.—That a train of waves emitted by a source of homogeneous light furnishes for length measurements a scale of extraordinary precision and constancy is now more or less familiar to every one, and the importance of such fundamental standards to spectroscopy, metrology, and precision optics generally is fully appreciated. For practical use very narrow (homogeneous) spectral lines must be used, and their wave-lengths should be carefully defined with respect to the conditions producing or influencing them. Several successive attempts have been made to define such a system of Standard Wave-lengths, each aiming at higher precision by making use of more refined methods of measurement or of more satisfactory light sources. The Solar Spectrum Standards first determined by Ångström in 1868, improved and extended by Rowland in 1888, have already been mentioned. The absolute measurements made by Michelson in 1893 on cadmium lines in terms of the standard metre at Paris showed that Rowland's wave-lengths were too large by 1/30,000, but spectroscopists paid little attention to this fact since there was still no reason to doubt the correctness in relative value of Rowland's numbers. In 1901, Fabry and Perot measured about 30 lines in the solar spectrum between 4643 and 6471 A by means of their interferometer, and their work revealed the fact that systematic errors existed in Rowland's Table, making the errors in relative magnitudes nearly 1 part in 100,000. The probable sources of those errors are discussed under the theory and use of diffraction gratings for wave-length measurements.

The question of standard wave-lengths was discussed at the first meeting of the International Union for Co-operation in Solar Research (21), and the necessity for establishing a new system of standard wave-lengths from artificial sources was agreed upon. At the second meeting of this Union in 1905, the following decisions (22) were reached: (1) The wave-length of a properly chosen radiation shall be taken as a primary standard of wave-lengths. The number which represents this wave-length shall be fixed one time for all; it shall define the unit of wave-length, which must differ as little as possible from 10^{-10} metre, and shall be called Ångström. (2) There is reason to choose secondary standards separated by not more than 50 Ångströms. These secondary standards shall be referred to the primary standard by a method of interference. The luminous source shall be an electric arc of 6 to 10 amperes.

In 1907 a report of the measurement by Benoit, Fabry, and Perot (22) of the wave-length of the red radiation from cadmium was presented at the third meeting of the International Union, and the following resolution was adopted (*loc. cit.* p. 20): "The wave-length of the red ray of light from cadmium, produced by a tube with electrodes, is

6438·4696 Ångström in dry air at 15° C. on the hydrogen thermometer, at a pressure of 760 mm. of mercury, the value of g being 980·67 (45°). This number will be the definition of the unit of wave-length."

The first determinations of secondary standards were reported in 1907 by Fabry and Buisson (*loc. cit.* p. 138), who measured the wave-lengths of 115 lines in the iron arc from 2373 to 6494 A by their interferometer method. Similar measurements were made by Eversheim (24) and by Pfund (25), and the mean values of the three independent determinations were adopted as secondary standards by the International Union in 1910, comprising 54 lines between 4282 and 6494 A.

The agreement between interferometer measurements of these wave-lengths is remarkably good (·002 or ·003 A), and there is reason to believe that the accepted secondary standards have a precision of 0·001 A, *i.e.* they are correct and reproducible to part 1 in several millions. This is the extreme limit which it is possible to obtain, since the most homogeneous lines of iron have a width of about 0·060 A, and a precision of 0·001 A corresponds therefore to 1/60 of the width of the line. Some of the differences between values of different observers can probably be accounted for by different arc conditions employed in different laboratories, and special investigations on the effect of arc conditions show that the wave-lengths of certain iron lines which are sensitive to pressure are also affected by current strength, length of arc, and portion of arc used. Recognition of these influences led, in 1913, to the adoption of more precise specifications of the iron arc in air as a source of international standards (26) as follows:—(1) Length of arc 6 mm. (2) Current of 6 amperes for wave-lengths greater than 4000 A, and 4 amperes or less for wave-lengths less than 4000 A. (3) Use direct current with positive pole above the negative and a potential of 220 volts; iron rods of 7 mm. diameter. (4) As a source of light use an axial part of about 2 mm. in the middle of the arc. (5) Use only iron lines of groups a, b, c, d (Mt. Wilson Classification).

It was also proposed, in 1910, that "the measurement of standards of the second order shall be extended to shorter and longer wave-lengths, and the arithmetical mean of three independent determinations shall be adopted as secondary standards." In the past ten years very little progress has been made in extending this system of secondary standards. The most extensive measurements were made by Burns, who compared directly with the cadmium primary standard the wave-lengths of 125 iron lines (27) from 5434 to 8824 A and 100 lines (28) from 2851 to 3701 A.

The wave-lengths which have been adopted as International Secondary Standards are collected in the following table:

International Secondary Standards.

3370·789	4375·934	5405·780
3399·337	4427·314	5434·527
3445·154	4466·556	5455·614
3485·345	4494·572	5497·522
3513·821	4531·155	5506·784
3556·881	4547·853	5569·633
3606·682	4592·658	5586·772
3640·392	4602·947	5615·661
3676·313	4647·439	5658·836
3677·629	4691·417	5709·396
3724·380	4707·288	5763·013
3753·615	4736·786	5857·759 Ni
3805·346	4789·657	5892·882 Ni
3843·261	4859·758	6027·059
3850·820	4878·225	6065·492
3865·527	4903·325	6137·701
3906·482	4919·007	6191·568
3907·937	4966·104	6230·734
3935·818	5001·881	6265·145
3977·746	5012·073	6318·028
4021·872	5049 827	6335·341
4076·642	5083·344	6393·612
4118·552	5110·415	6430·859
4134·685	5167·492	6494·993
4147·676	5192·363	6546·250
4191·443	5232·957	6592·928
4233·615	5266·569	6678·004
4282·408	5302·315	6750·163
4315·089	5324·196	..
4352·741	5371·495	..

In addition to the above determinations of secondary standards, many other measurements of wave-lengths in the iron arc spectrum have been made, some with interferometers and others with diffraction gratings, partly to set up a system of tertiary standards and partly to investigate further the effect of the operating conditions of the arc upon its wave-lengths. Among such investigations the most important in some respects appears to be that of the so-called "pole effect" on arc lines. This has been studied in some detail at the Mt. Wilson Observatory (29), where it is found that certain iron lines which are also sensitive to pressure give wave-lengths in the centre of the arc slightly different from those near the negative pole (one or two millionths), and also that this effect is somewhat reduced when, instead of using two iron electrodes, a carbon electrode and an iron one are used (30).

Corrections to wave-lengths measured in air also deserve mention in connection with the exact measurements. The Rowland system

of standards was defined in air at 20° C. and 760 mm. of mercury, and correction tables to reduce such wave-lengths or oscillation frequencies to vacuum values are given in Kayser's *Handbuch*, ii. 514. Since the new International System specifying wave-lengths at 15° C. has come into general use, there has been no uniformity in such corrections. The resulting confusion has been pointed out by Birge (31), who recommends that the recent work of the Bureau of Standards (18) on the index of refraction of air for wave-lengths from 2218 to 9000 A be referred to for these corrections in all future work. International adoption and use of the new tables will exclude all further possibility of ambiguity or question about the conversion of wave-lengths to standard air conditions or to vacuum when such wave-lengths are not actually observed under these conditions.

International co-operation ceased during the war, but a new International Astronomical Union, representing several of the allied countries, was organised in Brussels in July 1919. A new committee on Standard Wave-lengths and Solar Spectrum Tables was appointed, and this committee will, no doubt, consider the possibilities of still further improving the precision of standards and methods of measuring wave-lengths.

§ (5) WAVE-LENGTH MEASUREMENTS WITH PRISM SPECTROGRAPHS.—Diffraction gratings and interferometers are very wasteful of light, so that the prism spectrograph must be employed for spectral analysis of faint sources such as the stars, planets, low voltage arcs, etc. Wave-length measurements with prism spectrographs are generally obtained by methods of interpolation after the positions of the unknown lines are measured relative to the positions of certain well-known lines whose wave-lengths have already been accurately determined. Either graphical interpolation or an interpolation formula may be used. The graphical method requires drawing the dispersion curve of the spectroscope, *i.e.* the curve showing the relation between wave-length and corresponding deviation. This curve is obtained by measuring the deviations of well-known lines in different parts of the spectrum and then plotting these numbers against the wave-lengths on cross-ruled paper. A smooth curve traced through these points is the calibration curve of the instrument. To determine the wave-length of an unknown line it is only necessary to measure its deviation, after which the wave-length can be read directly from the curve.

Hartmann's interpolation formula is in common use, and very satisfactory for the purpose of determining wave-lengths of lines in prismatic spectra. It states the relation between wave-length λ and refractive index μ as follows:

$$\lambda = \lambda_0 + \frac{c}{(\mu - \mu_0)^{1/a}},$$

where λ_0, c, μ_0, and a are constants. For relatively short ranges of wave-length, a, which has a value of about 1·2, may be placed equal to unity, and μ may be replaced by the linear distance n between lines along the focal curve. Then

$$\lambda = \lambda_0 + \frac{c}{n - n_0}.$$

Hartmann pointed out that λ_0 is a constant for any particular spectroscope and may be determined once for all, while n_0 represents some definite point on the linear scale. The known wave-lengths λ_1, λ_2, and λ_3 and corresponding scale readings n_1, n_2, and n_3 for three lines are sufficient to calculate the three constants λ_0, c and n_0, as follows:

$$\lambda_0 = \lambda_3 + \frac{\lambda_1'(n_1 - n_2)}{S(n_3 - n_2) + (n_1 - n_2)},$$

$$c = (\lambda_0 - \lambda_3)(n_0 - n_3),$$

$$n_0 = \frac{(\lambda_0 - \lambda_3)(n_3 - n_2) + \lambda_2' n_2}{\lambda_2'},$$

where

$$\lambda_1' = \lambda_1 - \lambda_3, \quad \lambda_2' = \lambda_2 - \lambda_3, \quad \text{and} \quad S = \frac{\lambda_1' - \lambda_2'}{\lambda_2'}.$$

The accuracy obtainable in wave-length measurements from prism spectrograms depends principally on the dispersion, and under the best conditions the limit is about 0·01 A. Spectroscopic work with prisms is, of course, limited to the range of wave-lengths for which transparent prism materials are available. The lower limit is set by fluorite, which transmits to 1200 A, and sylvine marks the upper limit at about 23 μ or 230000 A.

§ (6) WAVE-LENGTH MEASUREMENTS IN THE SCHUMANN REGION.—The most precise determinations of light waves exist for the range from about 2000 to 9000 A, that is, in the region of the spectrum which is most easily photographed. It is very difficult to work with shorter waves because of the strong absorption of their energy by the air and by the gelatine of the photographic emulsions. Schumann (32) first overcame these difficulties with a vacuum spectrograph, fluorite prism, and special photographic plates. Consequently, these short waves are commonly called Schumann waves.

The first successful attempt at accuracy in length measurements of Schumann waves was made by Lyman (33). The wave-lengths of the hydrogen lines in this Schumann region were measured to serve as standards. The method used is briefly as follows:

Two illuminated slits S and S′, placed on a circle whose diameter is the grating's radius of curvature, will give images at I and I′, and spectra corresponding to each (*Fig.* 9). The two first-order spectra will be shifted with respect to each other by an amount depending on the distance between the slits. At a given point on a photographic plate between I and I′ the light brought to a focus from S will be of shorter wave-length than that from S′. The amount of this shift was determined by comparing known lines of the aluminium and iron spectra at one end of the plate. Then the longer wave portion of the iron spectrum (3100 to 2400 A) was superposed on the shorter wave portion (1900 to 1200 A) of the hydrogen spectrum and the wave-lengths in the former reduced by the shift. The accuracy in this region does not surpass 0·2 or 0·3 A.

The shortest waves observed by Schumann himself were about 1200 A. Lyman extended the spectrum down to about 600 A, and has collected the important tables of wave-lengths and other optical data relating to the Schumann region in a monograph entitled *Spectroscopy of the Extreme Ultra-violet.* Very recently Millikan has pushed the wave-length limit an octave lower, so that a gap of less than five octaves remains between the shortest ultra-violet waves and the longest X waves thus far observed.

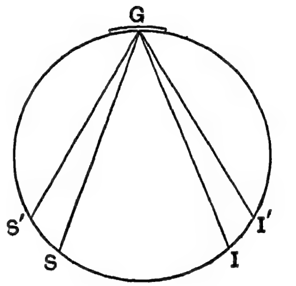


Fig. 9.

§ (7) Wave-length Measurements in the Infra-red.—Beyond the photographic limit, investigation of the infra-red spectrum by means of the heating effect of the rays has been carried on with the aid, chiefly, of thermopile, radiometer, radio-micrometer, and bolometer. The thermopile consists of a number of junctions of dissimilar metals, *e.g.* bismuth and antimony, iron and constantin. If radiation falls on alternate junctions a galvanometer in the circuit will give a deflection proportional to the heating effect.

The radiometer is a modification of the instrument devised by Crookes in which mica vanes, accurately mounted on a central spindle *in vacuo*, rotate when placed in the path of radiant energy. The radio-micrometer[1] invented by Boys and d'Arsonval simultaneously is a type of galvanometer. A single thermocouple is suspended by a quartz fibre with the plane of the couple in the line joining the poles of a magnet. When radiation falls on a junction the current generated causes the couple to turn in the magnetic field. In 1881 Langley (34) announced his actinic balance or bolometer,[2] which measures the temperature of a very fine strip of metal in terms of the change of electrical resistance. Two similar strips are placed in two of the arms of a Wheatstone bridge. When radiation falls on one of them the resistance changes and causes a deflection of the galvanometer.

In connection with this method of spectrum investigation the name of Langley (35) became famous. His bolometer was an exceedingly great advance upon any of the apparatus which had previously been constructed. With a bolometer strip 0·5 mm. broad and ·002 mm. thick he reached a sensitiveness of 10^{-6}° C. With such instruments he investigated the heat spectrum of the sun and the absorption of the earth's atmosphere. In his final work on the solar spectrum Langley used a fixed-arm type of spectrometer, with lenses and prisms of rocksalt, and the spectrum was made to pass over the bolometer strip by rotating the prism. The galvanometer readings were recorded on photographically sensitive paper whose motion was co-ordinated with that of the prism. From the known dispersion of the prism the wave-length of any spectrum line shown upon the record could be found, and also, from the length of the galvanometer throw, its intensity estimated. Langley thus mapped the solar spectrum as far as 5·5 μ, and observed 700 lines between A (·7604 μ) and this limit.

For more than a decade Langley was practically the only worker in this field. In 1892 Rubens (36) began to measure the dispersion of various substances for long wave-lengths. With a bolometer he measured the dispersion of crown and flint glasses, water, and various liquid hydrocarbons to 3 μ, and of quartz, rocksalt, and fluorite to 3·5 μ. Later he extended the work on fluorite to 6·48 μ.

In 1894 Paschen (37) went out to 9·3 μ by using a fluorite prism, and Rubens (38) and Trowbridge in 1897 pressed forward to 23 μ by using prisms of rocksalt and sylvine. The continuous spectrum was thus extended to thirty-nine times the wave-length of sodium yellow, or five octaves into the infra-red. A further extension in this manner seemed impossible, because no substances were known which were more transparent than sylvine to the long waves. The extreme limit of prism

[1] See "Radio-micrometer and other Instruments," Vol. III.

[2] See "Radiant Heat, Instruments for the Measurement of," § (18).

transmission is fixed by sylvine at 23 μ. With rocksalt and fluorite the limits are already reached at 15 μ and 9 μ respectively. It is an interesting fact that the absorption which makes it impracticable to use these substances further as prism material furnished the key to still greater penetration into the infra-red region.

The presence of these absorption bands is attributed to a resonance phenomenon which

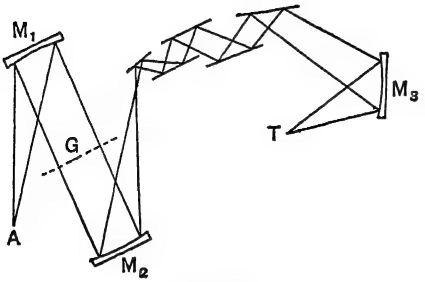


FIG. 10.

causes a large part of the incident radiation to be re-emitted, so that the particular substance shows metallic (*i.e.* high) reflection at these points. This fact is the basis of the method of isolating long waves which was worked out by Rubens (39) and Nichols in 1899. If the radiation of a source falls upon a plate of rocksalt, for example, all wave-lengths are transmitted except in absorption regions; the wave-lengths corresponding to the maximum absorption are almost wholly reflected. The diagram (*Fig.* 10) of Rubens' apparatus shows light from an incandescent Welsbach burner A passing through a wire grating G and undergoing several successive reflections before being focussed on thermopile T. The final radiations thus obtained by successive reflections from a substance were called "Reststrahlen" or residual rays, and possess the greatest energy for almost the same wave-lengths as the maxima of the absorption bands for the substance. For example, the absorption bands computed by means of the Ketteler-Helmholtz dispersion formula are at 35 μ for fluorite, 56 μ for rocksalt, and 67 μ for sylvine. Rubens and Nichols found experimentally for fluorite 24 μ and 34 μ, for rocksalt 52 μ, and for sylvine 63·4 μ; very good agreement if one considers that the dispersion formula requires the reflection maximum at somewhat shorter wave-length than the absorption. It was suspected that for other substances, especially those with larger molecular weight, still longer waves could be found. This was confirmed by Rubens, and is shown in the following table:

Substance.	Mean λ of Reststrahlen.	Molecular Weight.
Rocksalt (NaCl) . . .	52·0 μ	58·5
Sylvine (KCl) . . .	63·4	74·6
Silver chloride (AgCl) .	81·5	143·4
Potassium bromide (KBr)	82·6	119·0
Thallium chloride (TlCl)	91·6	239·5
Potassium iodide (KI) .	94·1	166·0
Silver bromide (AgBr) .	112·7	188·0
Thallium bromide (TlBr)	117·0	284·0
Thallium iodide (TlI) .	151·8	331·0

The infra-red spectrum was thus pushed to nearly 8 octaves above the visible. Because of the limited energy of sources, further extension to longer waves by the Reststrahlen method is difficult in spite of the refinement of modern apparatus.

A still further improvement in the method of isolating very long waves was made by Rubens (40) and Wood in 1911. The method is based on the selective refraction and selective absorption of quartz. Quartz has an absorption band beginning at 4·5 μ which reaches far out into the infra-red, but for wave-lengths longer than 70 μ it becomes transparent.

At the same time, in consequence of its anomalous dispersion, quartz has a very large index of refraction for 70 μ; at 63 μ, $n = 2{\cdot}19$ and approaches 2·14 with longer wave-lengths. For visible and short heat waves its index of refraction ranges from 1·43 to 1·55. A quartz lens, therefore, has a much shorter focal

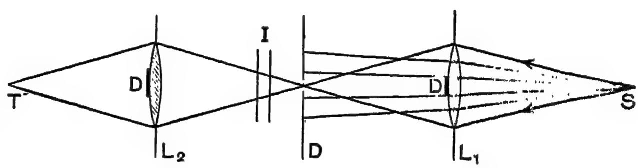


FIG. 11.

length for extremely long waves than for the shorter ones. This principle was used by Rubens and Wood in their apparatus, *Fig.* 11, in which the long waves are selected from the source S by means of the quartz lenses L_1 and L_2 and opaque screens D which stop the short waves. The long waves pass through an interferometer I and are focussed on a thermopile T. The great advantage of this method of focal isolation is that large angular apertures can be used and the radiation is weakened only by reflection and absorption by the two quartz lenses. Its only disadvantage is that all the radiations for which n approximates 2·14 pass through. The wave-lengths are determined

TABLE OF WAVE-LENGTHS

Waves.	Octave No.	Frequency × 10^{14}.	Wave-length λ μ.	Remarks.
Unknown ..	−18	1,000,000	0·000003	
	−17	491,520	0·000006	
γ and X	−16	245,760	0·000012	−·000007 μ. Rutherford, 1914. γ from radium B
	−15	122,880	0·000024	
	−14	6,1440	0·000048	—·0000292 X_β from Neodymium
	−13	30,720	0·000098	−·0000560 μ. X_α from silver. Moseley, 1914
	−12	15,360	0·000195	−·0001365 μ. γ from radium B. Rutherford, 1914
	−11	7,680	0·00039	
	−10	3,840	0·00078	
	− 9	1,920	0·00156	−·0008364 μ. X_α from aluminium. Moseley, 1914 −·0012346 μ. X_α from zinc. Friman, 1916
Unknown ..	− 8	960·0	0·0031	
	− 7	480·0	0·006	
	− 6	240·0	0·012	
	− 5	120·0	0·025	
Schumann..	− 4	60·0	0·05	−·036 μ. Millikan, 1920
	− 3	30·0	0·1	−·06 μ. Lyman, 1914
	− 2	15·0	0·2	−·12 μ. Lower limit of fluorite transparency. Schumann. 1893 −·18 μ. Lower limit of air transparency, and of quartz, rock-salt, and photographic gelatine
Ultra-violet	− 1	7·50	0·4	−·35 μ. Lower limit of glass transparency ·4 μ. Visible limit in violet
Visible. ..	0	3·75	0·8	−·64384696 μ. Fundamental standard wave-length ·8 μ. Visible limit in red
Infra-red .	+ 1	1·875	1·6	−1 μ. Upper limit of photography
	+ 2	0·940	3·2	−3 μ. Upper limit of glass transparency
	+ 3	0·470	6·4	−5·3 μ. Langley, 1886

TABLE OF WAVE-LENGTHS (*continued*)

Waves.	Octave No.	Frequency $\times 10^{14}$.	Wave-length λ μ.	Remarks.
Infra-red (*continued*)	+ 4	0·235	12·8	-9 μ. Upper limit of fluorite transparency
	+ 5	0·118	25·6	-15 μ. Upper limit of rock-salt transparency -23 μ. Upper limit of sylvine transparency
	+ 6	0·059	51·2	
	+ 7	0·029	102·4	-52 μ NaCl -63 μ KCl -82 μ KBr -113 μ AgBr — Reststrahlen. Rubens and co-workers, 1897–1914
	+ 8	0·014	204·8	-152 μ Tl I
	+ 9	0·007	409·6	-218 μ, -342 μ Quartz-lens focal isolation of Hg arc emission bands. Rubens and von Baeyer, 1911
Unknown	+10	0·0035	819·2	
	+11	0·0018	1638·4	
Electric	+12	0·0009	3276·8	-2000 μ. O. von Baeyer, 1911
	+13	0·00045	6553·6	-4000 μ. Lampa, 1897 -6000 μ. Lebedew, 1895
	+14	0·00022	13107·2	
	+15	0·00011	26214·4	
	+16	0·00005	52428·8	
	+17	0·00003	104857·6	-100000 μ. Righi, 1893
	+18	0·000015	209,715	
	+19	0·000007	419,430	
	+20	0·000003	838,860	$-660,000$ μ. Hertz, 1888
	+21	0·0000016	1,677,720	
	+22	0·0000008	3,355,440	
	+23	0·0000004	6,710,880	
	+24	0·0000002	13,421,760	
	+25	0·0000001	26,843,520	
	+26	0·00000005	53,687,040	$-$Range of wave-lengths used in radio communication in 1920, 50 m. to 50,000 m.
	+27	0·000000025	107,374,080	Longest electric waves generated about 1,000,000 km.

from "visibility" curves obtained by plotting increasing separation of the interferometer plates as abscissae and corresponding energy readings as ordinates. In this way it was found that the radiation from a Welsbach incandescent mantle contains waves over 150 μ long, the maximum radiation being about 100 μ. Rubens (41) and von Baeyer in 1911 investigated by this method the long waves from a quartz mercury lamp and found an emission region between 200 and 400 μ with maxima at 218 μ and 342 μ. Here is a light wave 1/3 mm. in length, or over 500 times the length of the yellow sodium ray. This is the longest wave isolated to date, and is only 2¼ octaves from the shortest electric waves ($\lambda = 2$ mm.) obtained by O. von Baeyer. The rapid progress in the knowledge of the region between the visible spectrum and electric waves in the last 30 years justifies the hope that the remaining gap will soon be filled up. An effort in this direction was made in 1914 by Rubens (42), and Schwarzschild, who tried to find solar radiation between 400 and 600 μ by the quartz lens method. Calculation predicted that radiation could be measured if the sun were assumed a black body at 6000° C. and if the atmosphere were transparent. However, not a trace of deflection could be observed. The explanation is probably that the water-vapour in the earth's atmosphere, whose absorption spectrum is complex and full of bands below 400 μ, also possesses strong absorption for still longer waves.

Standards of wave-length in the infra-red part of the spectrum have been determined with considerable accuracy by Paschen (43) and by Randall (44), using thermopiles with which to explore the normal spectrum of concave gratings. Their investigations cover the arc spectra of a large number of the chemical elements. The measurements extend to wave-lengths of 30 to 40 μ, and are probably not more than a few tenths of an Ångström in error.

§ (8) Table of Wave-lengths.—The charts on pages 893-894 show the entire range of wave-lengths of electromagnetic waves which are within the domain of scientific investigation at the present time.

References

1. Young, *Roy. Soc. Phil. Trans.*, 1802, ii. 12-48.
2. Wollaston, *Roy. Soc. Phil. Trans.*, 1802, ii. 365-380.
3. Fraunhofer, *Gilbert's Ann.*, 1817, lvi. 264-313.
4. Fraunhofer, *Denkschr. d. K. Akad. d. Wiss. München*, 1821, viii. 1-76.
5. Bell, *Phil. Mag.*, 1887, xxiii. 265, and 1888, xxv. 255, 360.
6. Rowland, *Phil. Mag.*, 1887, (5), xxiii. 257, and 1889, xxvii. 479.
7. Runge and Paschen, *Wied. Ann.*, 1897, lxi. 641.
8. Kayser, *Astroph. Journ*, 1904, xix. 157.
9. Howell, *Astroph. Journ.*, 1914, xxxix. 230.
10. Jewell, *Trans. I. U.S. R.*, 1906, i. 42.
11. Goos, *Zeits. Wiss. Phot.*, 1911, ix. 173.
12. Birge, *Astroph. Journ.*, 1916, xliii. 81.
13. Lyman, *Phys. Rev.*, 1903, xvi. 257.
14. Michelson, *Trav. et Mém.*, 1893. xi.
15. Benoit, Fabry, and Perot, *Trav. et Mém.*, 1913, xv.
16. Fabry and Perot, *Ann. de chimie et de physique*, 1902, xxv. 110.
17. Meggers, *Bureau of Standards, Bull.*, 1915, xii. 203.
18. Meggers and Peters, *Bureau of Standards, Bull.*, 1918, xiv. 697.
19. Meggers, *Bureau of Standards, Bull.*, 1915, xii. 199.
20. Fabry and Perot, *Astroph. Journ.*, 1902, xv. 73 and 261
21. *Astroph. Journ.*, 1914, xx. 313
22. *Trans. I. U.S. R.*, 1906, i. 234.
23. *Trans. I. U.S. R.*, 1907, ii. 109.
24. Eversheim, *Ann. d. Physik*, 1911, (4), xxxvi. 1071.
25. Pfund, *Astroph. Journ.*, 1908, xxviii. 197.
26. *Trans. I. U.S. R.*, 1914, iv. 58.
27. Burns, *Journ. de physique*, 1913, (5), iii. 457.
28. Burns, *Bureau of Standards, Bull.*, 1915, xii. 179.
29. St. John and Babcock, *Astroph. Journ.*, 1915, xlii. 231.
30. St. John and Babcock, *Astroph. Journ.*, 1917, xlvi. 138.
31. Birge, *Astroph. Journ.*, 1918, l. 72.
32. Schumann, *Ber. Wien. Akad.*, 1893, cii. II.*a*, 415 and 627.
33. Lyman, *Astroph. Journ.*, 1906, xxiii. 181.
34. Langley, *Am. Sci.*, 1881, (3), xxi. 187.
35. Langley, *Wied. Ann.*, 1884, xxii. 598; *Phil. Mag.*, 1888, (1), xxvi. 505.
36. Rubens, *Wied. Ann.*, 1891, xlv. 238.
37. Paschen, *Wied. Ann.*, 1894, liii. 301.
38. Rubens and Trowbridge, *Wied. Ann.*, 1897, lx. 724.
39. Rubens and Nichols, *Wied. Ann.*, 1897, lx. 418.
40. Rubens and Wood, *Preuss. Akad. Wiss. Berlin, Ber.*, 1910, 1122.
41. Rubens and von Baeyer, *Preuss. Akad. Wiss. Berlin, Ber.*, 1911, 339 and 666.
42. Rubens and Schwarzschild, *Preuss. Akad. Wiss. Berlin, Ber.*, 1914, 702.
43. Paschen, *Ann. d. Physik* (4), xxvii. 537; xxix. 625; xxxiii. 717.
44. Randall, *Astroph. Journ.*, 1911, xxxiv. 1, and 1919, xlix. 42 and 54.

Wave-lengths of Light, first determinations of, corresponding to the different colours, made by Thomas Young. See "Wave-lengths, The Measurement of," § (1).
 Table of. See *ibid.* § (8).

Wave Surface, Fresnel's: the surface over which the energy emitted at a given instant from a point source in a biaxial crystal is distributed after any given interval of time; it is the envelope of the series of planes obtained by drawing, at any time, the wave-fronts. See "Light, Double Refraction of"; also "Polarised Light and its Applications," §§ (6) and (22).

Weber Illumination Photometer. See "Photometry and Illumination," § (58).

Whitman Photometer: a form of flicker photometer. See "Photometry and Illumination," § (96).

Wien's Displacement Law. See "Radiation Theory," § (5).

Wild Photometer. See "Photometry and Illumination," § (99); also (spectrophotometer) "Spectrophotometry," § (12).

Wild's Polarimeter. See "Polarimetry," § (8).

— X —

— Y —

— Z —

INDEX

(Numbers refer to pages)

END OF VOL. IV

Printed in Great Britain by R. & R. Clark, Limited, *Edinburgh.*